Extracting information from trading volume

Dominique Y. Dupont †

April 2002

Abstract:
This paper introduces a method to extract information about the true value of a traded asset using the market price and the equilibrium trading volume when the joint distribution of the traders' demands, the market price, and the asset value is known and the market clears. The paper applies the method to a chosen noisy rational-expectations model.
We first condition only on the trading volume and then on volume and price. Results from the multivariate conditioning contrast with those of the univariate conditioning. Conditioning on volume alone seems to yield little information on the true value of the asset. Still, the volume-based conditional distribution of the true asset value has slightly fatter tails than the unconditional distribution when volume is high, and thinner tails when volume is low.
In contrast, in the multivariate conditioning, when the market price is above its mean, a surge in volume can lower the conditional likelihood in the upper tail. Indeed, the conditional value of the traded asset can be decreasing in trading volume for a given price. It can even be decreasing in price when volume is high. Finally, we apply the method to a market-making context using the Kyle (1985) model.

JEL Classification: G12, G13, C63. Key words: Trading volume, information

---

†University of Twente, School of Management and Technology, Department of Finance and Accounting. Postbus 217, 7500 AE, Enschede, The Netherlands. Tel: +31 53 489 44 70, Fax: +31 53 489 2159, Email: d.dupont@sms.utwente.nl. Formerly at Eurandom-TUE, P.O. Box 513 - 5600 MB Eindhoven. The author thanks seminar participants at the Federal Reserve Board, Erasmus University, ESSEC, the 1998 French Finance Association Meeting, and at Eurandom, in particular Johan Segers. The usual disclaimer applies.

Introduction

One function of financial markets in any economy is to aggregate the information that agents might have on the traded asset and to disseminate this information throughout the economy. Economists have studied extensively the role the market price plays in aggregating and relaying information and practitioners use market prices to guide their economic decision making. However, price is not the only informational output of markets. Other indicators of market activity, such as trading volume, are available to the general public. For example, data on daily trading volume in futures markets are freely available on the websites of the major exchanges and in the financial press. Hence, it may be interesting to develop a method to extract information on the true value of the traded asset based on the trading volume and the market price.

This paper introduces such a method when the joint distribution of the traders' demands, the market price, and the asset value is known and the market clears. The paper also applies the method to a chosen noisy rational-expectations model and a market-making model.

We first derive a general result: if a random variable is symmetrically distributed around zero with the traders' demands, the covariance between this random variable and the trading volume is zero (Proposition 1). Proposition 4 extends this result by showing that the conditional distribution of such a random variable conditioned on the market price and the trading volume is symmetric and consequently its conditional expectation based on volume is zero. This result is quite intuitive. For example, as trading volume is defined as the sum of the buy orders, a surge in volume indicates a large flow of executed buy orders. But as the market clears, this strong influx of executed buy orders corresponds necessarily to a equally strong flow of executed sell orders. Hence, intuitively, trading volume should offer little guidance about the mean of the asset value.

Under the additional assumption that the random variable and the traders' demands are normally distributed, and that there are only three traders, we obtain a closed-form expression for the covariance between the absolute value of this random variable and volume and show that this covariance is positive unless the variable is independent of the traders' demands (Proposition 2). The random variable under consideration may be the true value of the asset, the price or, in a dynamic model, the price change.

The results presented in this paper can be extended to market-making environments if the same distributional assumptions are maintained. Any market-marking model can be reframed as a market-clearing model by considering the market-maker as one of the traders. Indeed, when there is one market maker and n traders, the trading volume is defined as $z = \frac{1}{2}\Sigma_{i=1}^{n}|y_i| + \frac{1}{2}|\Sigma_{i=1}^{n} y_i|$, where y_i is the trader i's demand. By defining the market maker's demand as $y_{n+1} = -\Sigma_{i=1}^{n} y_i$, and using the fact that $y_i = y_i^+ + y_i^-$ and $|y_i| = y_i^+ - y_i^-$, one gets $z = \frac{1}{2}\Sigma_{i=1}^{n+1}|y_i| = \Sigma_{i=1}^{n+1} y_i^+$, which is the definition of volume used in the paper.

Empirical observations seem to support the predictions of the paper concerning the covariance between trading volume and the level or the absolute value of any random variable whose joint distribution with the traders' demand is symmetric or normal. In almost all empirical studies, the covariance between the absolute value of the price change and trading volume is positive, while the covariance between the price change per se and volume is statistically zero, at least on futures markets, where the absence of short-sale constraints preserves the symmetry between buy and sell orders.

We also derive the conditional distribution of the true value of the asset conditioned on trading volume alone or on price and trading volume and present quantitative results based on a chosen noisy rational expectation model. Given the joint distribution of the equilibrium price (p), the traders' demands, and the true value of the asset (x), and letting z be the trading volume, we can compute $cov(|p|, z)$, $cov(|x|, z)$, and the conditional density of x conditioned on p and z or on z alone.

Results from the multivariate conditioning contrast with those of the univariate conditioning. Conditioning on volume alone seems to yield little information on the true value of the asset: $E[x|z] = E[x]$ and the conditional density of x given z differs little from the unconditional density (except for very large values of z). Still, the volume-based conditional distribution of the true asset value has slightly fatter tails than the unconditional distribution when volume is high, and thinner tails when volume is low. In contrast, in the multivariate conditioning, when the market price is above its mean, a surge in volume can lower the conditional likelihood in the upper tail. Indeed, the conditional value of the traded asset $E[x|p, z]$ can be decreasing in trading volume for a given price. It can even be decreasing in price when volume is high.

Results obtained in the Kyle market-making framework differ from those obtained in the market-clearing framework. In Kyle (1985), the price fixed by the dealer equals the conditional expectation of the true asset value conditioned on the order flow (the sum of the informed trader's and the liquidity trader's buy or sell orders). This may explain why $E[x|p, z] = E[x|p]$ with probability 1 in this framework. However, including volume in the conditioning set gives additional information about the tails of the distribution of x.

In the remainder of the paper, section 1 briefly reviews the literature, section 2 presents results on the covariance between trading volume and the level or the absolute value of any random variable. Section 3 shows how to condition on trading volume or on trading volume and price. Finally, section 4 applies the method developed in section 3 to a noisy rational expectations model and section 5 applies it to a Kyle (1985) market making model.

1 Review of the empirical literature

1.1 Theoretical literature

The results of this paper apply to the majority of the existing models. Microstructures models often use normally distributed random variables. This is the case for, among others, Diamond and Verrechia (1981), Kyle (1985), Wang (1992), (1994), Foster and Viswanathan (1994), and He and Wang (1995). For them, Propositions 1 to 4 apply, independently of the particular assumptions of each model. The only necessary assumptions are that the random variable about which one wants to extract information and the traders' demands for the asset be normally distributed around zero (or, for weaker, more general results, that this variable be jointly symmetrically distributed with the traders' demands), and that the market clears. The traders' utility functions, their motives to trade and their rationality can be left unexamined. Furthermore, the symmetry argument behind Proposition 1 also applies to some non-normal distribution, like the elliptically contoured distributions used by Foster and Viswanathan (1993).

Moreover, the model developed by Blume et al (1994), which is not based on normal or elliptical distributions, also suggests a positive relation between absolute price change and trading volume. Blume et al construct a model where the precision of the signals observed by the traders is stochastic. Simulations based on their model show a positive relation between volume and the absolute value of price change, and a symmetric relation between price change per se and volume (i.e., a large volume is associated with a negative or a positive movement in the asset price). The V shape of the price-volume relation confirms the positive correlation between the magnitude of the price change and trading volume.

1.2 Empirical literature

In his extensive review of the literature, Karpoff (1987) points out that many empirical studies have found a positive correlation between the absolute value of price change and trading volume. Furthermore, Karpoff reports that the correlation between price change per se and volume is statistically insignificant for futures markets while it is positive and statistically significant for equity markets, a pattern he attributes to the existence of short sale constraints in stock and bond markets (see also Karpoff (1988)). Proposition 1, which implies a zero covariance between volume and price change per se, rests on the assumption that the traders' demands are symmetrically distributed. Short sale constraints would introduce asymmetries in the demands and could hence create a positive correlation between price change and volume, despite market clearing.

Since 1987, many economists have studied the relation between trading volume and the magnitude of the price change, and between volume and price change per se, on stock data and futures data. These empirical studies tend to confirm Karpoff's conclusions. Using hourly New York Stock Exchange data between 1979 and 1983, Jain and Joh (1988) find a significant positive relation between daily trading volume and the absolute value of the Standard & Poor 500 index returns. Gallant et al (1992) use daily New York Stock Exchange data between 1928 and 1987 to estimate the joint density of current price change and volume conditional on past price changes and volume. They find that "the direction of the daily change in the stock market is unrelated to contemporaneous volume," (p. 223) and that unusually high volumes are associated with large price changes.[1] Goodman (1996) uses a random sample of 50 stocks traded on the New York and the American Stock Exchanges between 1993 and 1994. His findings confirms that the absolute value of the price change is positively correlated with trading volume, and shows that strong volume is associated with extreme price movements, both positive and negative.

As for futures markets, Karpoff uses daily data on futures contracts for 9

[1] However, using daily Helsinki Stock Exchange data between 1977 and 1988, Martikainen et al (1994) find some positive and statistically significant cross-correlations between contemporaneous values of stock returns and volume.

commodities and 3 financial instruments (also called commodities thereafter) between January 1972 to December 1979. Fifty-one percent of the 442 contracts studied displayed a statistically significant positive relation between the absolute value of the price change and volume whereas only 6 percent of them showed a statistically significant positive relation between the price change per se and volume. The analysis is repeated on 12 time series (one for each commodity) constructed from the futures contracts data. The relation between the magnitude of the price change and volume is positive for all the commodities, and statistically significant for 9 of them. None of the commodities shows a significant relation between price change per se and volume.

Foster (1995) uses daily data on two oil futures contracts between January 1990 and June 1994, and one oil futures contract between January 1984 and June 1988. As in Gallant et al (1992), volume data are detrended and expressed in logarithms, and are first grouped in several classes per size. The relative price change is then plotted against the volume classes. The magnitude of the price change typically increases with volume but the direction of the change is not related to trading volume. This conclusion still holds when actual volume data are used instead of volume classes.

2 Co-moments with trading volume

Let y_i be trader's i demand for the asset ($i = 1, \ldots, n$) and z be the aggregate trading volume, that is, the sum of the buy orders, $z = \Sigma_{i=1}^n y_i^+$, with $y_i^+ = y_i \, I[y_i > 0]$ where $I[\,]$ is the indicator function. The asset market clears, hence $\Sigma_{i=1}^n y_i = 0$. Let $y = (y_i)_{i=1}^n$, y is assumed to have mean zero. The objective is to compute $cov(x, z)$ and $cov(|x|, z)$ where x stands here for any random variable although it will refer to the true asset value later on. Propositions 1 and 2 can hence also be applied to the market price, and the price change (in a dynamic framework).

Proposition 1. *Let x and y be a random variable and a n-dimensional vector, respectively, such that $\Sigma_{i=1}^n y_i = 0$ and (x, y) be symmetrically distributed around zero. Let $z = \Sigma_{i=1}^n y_i^+$, where $y_i^+ = I[y_i > 0] y_i$. Then, $cov(x, z) = 0$.*

The proof is in the Appendix. This proof does not depend on the value of n. However, when working with the distribution of z, we have to consider the sign of each y_i, $i = 1, \ldots, n$. One constraint imposed by the market-clearing condition is that not all y_i can be of the same sign, which implies that still $m = 2^n - 2$ cases have to be considered. In other words, we need to consider the sign of $y_1, \ldots, y_{n-1}$ and the sign of $\sum_{i=1}^{n-1} y_i$ when not all the y_i are of the same sign. For simplicity, assume that $n = 3$. We have then to consider $m = 6$ cases (see Table 1). Let $1 \leq j \leq m$ and A_j a cartesian product of n (adequately chosen) segments equal to $(-\infty, 0]$ or $(0, +\infty)$ so that the A_j form a partition of R^n. In each case, z can be written as the sum of two non-negative random variables that are linear combinations of the y_i, $i = 1, \ldots, n-1$, that is,

$$z = \sum_{j=1}^{m} I[y \in A_j] \, z = \sum_{i=j}^{m} I[v_j^1 > 0, v_j^2 > 0] \, (v_j^1 + v_j^2), \tag{1}$$

For example, when $y_1 > 0$, $y_2 \leq 0$ and $y_1 + y_2 > 0$, $z = y_1 = (-y_2) + (y_1 + y_2)$ with $-y_2 \geq 0$ and $y_1 + y_2 > 0$. Moreover, $y_1 > 0$, $y_2 \leq 0$ and $y_1 + y_2 > 0$, if and only if $-y_2 \geq 0$ and $y_1 + y_2 > 0$. Equation (1) makes it easy to handle trading volume.

Proposition 2. *Assume that $n = 3$ and that (x, y) is normally distributed with mean zero. Then,*

$$\begin{array}{lcl} E[z] & = & \frac{1}{\sqrt{2\pi}}(\sigma_{y_1} + \sigma_{y_2} + \sigma_{(y_1+y_2)}) \\ var(z) & = & 2\big(\sigma_{y_1}\sigma_{y_2}\, g(\rho_{y_1,y_2}) + \sigma_{y_1}\sigma_{y_1+y_2}\, g(-\rho_{y_1,y_1+y_2}) + \sigma_{y_2}\sigma_{y_1+y_2}\, g(-\rho_{y_2,y_1+y_2})\big) \\ & + & \frac{1}{2}\big(\sigma^2_{y_1} + \sigma^2_{y_2} + \sigma^2_{y_1+y_2} - \frac{1}{2\pi}(\sigma_{y_1} + \sigma_{y_2} + \sigma_{(y_1+y2)})^2\big) \\ cov(|x|, z) & = & \frac{\sigma_u}{\pi}\big(\sigma_{y_1}\, h(\rho_{u,y_1}) + \sigma_{y_2}\, h(\rho_{u,y_2}) + \sigma_{y_1+y_2}\, h(\rho_{u,y_1+y_2})\big) \end{array} \tag{2}$$

where σ and ρ are the standard deviation and the correlation coefficient, respectively. The functions g and h are defined as:

$$\begin{array}{lcl} g(t) & = & t\,(\frac{1}{4} + \frac{1}{2\pi}\arcsin(t)) + \frac{1}{2\pi}\sqrt{1-t^2} \\ h(t) & = & t\,\arcsin(t) + \sqrt{1-t^2} - 1 \end{array} \tag{3}$$

Equation (2) implies that $cov(|x|, z) \geq 0$, and $cov(|x|, z) = 0$ if and only if u and y are independent.[2]

The proof is in the Appendix.

3 Conditioning on trading volume and price or on trading volume alone

The aim is to compute the conditional density of x on z or on (p, z). To do so, we first need to establish the existence of the density of (x, p, z) and compute it. In the remainder of the paper, u, t and h denote the realizations of the random variables x, p, and z, respectively, and $f_z(h)$, $f_{p,z}(t, h)$, and $f_{x,p,z}(u, t, h)$ are the density functions of z, (p, z), and (x, p, z), respectively.

Proposition 3. *Let $f^i(x, p, v_i^1, v_i^2)$ and $f^i(p, v_i^1, v_i^2)$ the densities of (x, p, v_i^1, v_i^2) and (p, v_i^1, v_i^2), $i = 1, \ldots, m$. Then, $f_{x,p,z}$ and $f_{p,z}$ exist and*

$$\begin{array}{lcl} f_{x,p,z}(u, t, h) & = & \sum_{i=1}^m \int_{s=0}^h f^i(u, t, s, s-h)\, ds \\ f_{p,z}(t, h) & = & \sum_{i=1}^m \int_{s=0}^h f^i(t, s, s-h)\, ds \end{array}$$

The proof is in the Appendix.

Let $f_{x|p,z}(u, t, h)$ be the conditional density of x given p and z, and $f_{x|z}(u, h)$ be the conditional density of x given z. Naturally,

$$\begin{array}{l} f_{x|p,z}(u, t, h) = \frac{f_{x,p,z}(u,t,h)}{f_{p,z}(t,h)} \\ f_{x|z}(u, h) = \frac{f_{x,z}(u,h)}{f_z(h)} \end{array} \tag{4}$$

[2] For all $t \in [-1, 1]$, $h(t) \geq 0$ and $h(t) = 0$ if and only if $t = 0$. To see that h is strictly increasing on $[0, 1]$, let $m(s) = h(sin(s))$ with $s \in [0, \pi/2]$ and compute $m'(s)$.

Proposition 4. *The conditional density of x conditioned on p and z (or z alone) is symmetrical around zero. This implies that*

$$\begin{array}{lll} P(x > \lambda|p,z)(t,h) & = & P(x < -\lambda|p,z)(-t,h) \\ E[x|p,z](-t,h) & = & -E[x|p,z](t,h) \end{array} \tag{5}$$

$$\begin{array}{lll} P(x > \lambda|z)(h) & = & P(x < -\lambda|z)(h) \\ E[x|z](h) & = & 0 \end{array} \tag{6}$$

The proof is in the Appendix. Proposition 4 implies that $cov(x,p) = 0$.

4 Conditioning on price and volume in a noisy rational-expectations model

Using a particular setup to model the market, we can compute the conditional density of x conditioned on p and z or z alone. Including the price in the conditioning set seems quite natural because it is an observable variable. However, conditioning on trading volume alone sheds some light on the difference between univariate and multivariate conditioning. For example, whereas the (conditional) probability of x being in its upper tail is increasing in z when volume alone is used, this probability can be decreasing in z when $p > 0$. Moreover, whereas $E[x|p]$ is an linear, increasing function of p, for z fixed (above its mean), $E[x|p,z]$ is a non-linear, potentially decreasing function of p.

4.1 Presentation of the model

We use a static, competitive equilibrium framework with one liquidity and $n-1$ informed traders, each observing a signal correlated with the true value of the asset, with correlation coefficient ρ. In this framework, developed by Dupont (1998), the informed traders have CARA utility functions and the state variables (the signals, the noise, the true value of the asset) are jointly normally distributed, so that the informed traders' demands are linear in the market price and their signals.

We also assumed that the liquidity trader's demand is linear in price and the noise. Making the liquidity demand price sensitive is one contribution of this approach. Assuming that the traders have rational expectations, that is, that they use the equilibrium market price in their conditioning information, we derive the joint distribution of the equilibrium price (p), the traders' demands (y), and the true value of the asset (x). Based on this, we compute the conditional distribution of the x using z alone or using p and z. Before turning to the conditional distributions, we present results on the distribution of z and the correlations between x and z and between x and p.

4.2 Qualitative characteristics of the joint distribution of (x,p,z)

Figure 1 plots the mean and the standard deviation of trading volume as the quality of the information of the informed traders improves (the upper panel) and the probability density function of trading volume for $\rho = .2$ and $\rho = .5$

(the lower panel). Both the mean and the volatility of trading volume rise in response to an increase in ρ because the informed agents react more strongly to the realization of their private signals when those are better correlated with the true value of the asset. This more aggressive trading stance raises the average volume and makes it more volatile.

Looking at the density function reveals that, although the equilibrium price and all the demands are normally distributed, the distribution of volume is skewed, with the mean above the median. This reflects the impact on the distribution of outlayers: rare events when trading volume is particularly strong. This theoretical results is consistent with the use of the lognormal distribution to estimate the distribution of trading volume in empirical observations (see for instance Gallant et al (1992)). The skewness seems more pronounced for low levels of ρ. As seen in Figure 1, when $\rho = .2$, the density function rises sharply for low values of z but falls slowly in the upper tail. The difference between the mean and the median relative to the standard deviation is greater for the lower value of ρ.

4.3 Conditioning on trading volume only

Because of the random vector $(x, p, y_1, \ldots, y_n)$ is symmetrically distributed around 0, $E[x|z] = E[x] = 0$. This does not imply that no information can be extracted from trading volume. Since $cov(|x|, z) > 0$, one should expect that the conditional probability of x being in the left and right tails would increase as z increases. Figure 2 displays the conditional and unconditional probabilities that x be above a certain level (fixed equal to 0.2) as z increases. As z becomes higher than some threshold, the conditional probability becomes greater than the unconditional probability. $P(x > \lambda|z)(h)$ appears to be a convex function of h, the realization of z, so that the conditional probability picks up at an increasing rate as z rises. However, the size of the increase is small: even for extreme realizations of the trading volume, the conditional probability is only barely above the unconditional probability. For example, when $z = 3$, $P(x > .2|z)$ is about 0.05 percentage points above $P(x > .2)$ although $z = 3$ is a fairly extreme realization of the trading volume: the probability of z being above this level is around 1.25 percent.

This seems to suggest that trading volume is only loosely related to trading volume. This is confirmed by looking at the conditional distribution of x conditioned on z (see Figure 3). When $z = 3$, the conditional density of x given z is nearly indistinguishable from the unconditional density, and we chose a more extreme value for the trading volume, $z = 15$ to be able to show a clear effect on the conditional density of x given z. Both the upper and the lower tails of the distribution becomes fatter. However, to get this effect, we have to pick such a high realization of z that the probability of observing higher levels of the trading volume is of the order of about 1 in a million.

A simple way to visualize the dependence between the trading volume and true value of the asset or the market price for a wide array of precision of the informed traders' signals is to graph the correlation between z and $|x|$ for all possible values of ρ (the middle panel of Figure 3). Comparing $corr(|x|, z)$ and $corr(|p|, z)$ is very striking. The trading volume is nearly uncorrelated with the true value of the asset while it is strongly correlated with its market price. Furthermore, the correlation between $|p|$ and z is at its highest when the one between $|x|$ and z is at its lowest. When $\rho = 0$, the trading volume is perfectly correlated with the market price and totally uncorrelated with the true value of the asset. Naturally, the correlation between the market price and the true value of the asset is also zero since no trader has information about the true value of the asset. The correlation between the trading volume and the price is 1 because the demands of every trader and the price is linear in the noise ε (see proof in the Appendix). As ρ increases the correlation between $|p|$ and z first decreases then increases but always remains well above the correlation between $|x|$ and z. This suggests that trading volume contains a lot of noise, a view confirmed by looking at the correlation between z and the private signals and that between z and the noise. The correlation with $|G_1|$ (resp., with $|G_2|$) increases with the precision of the signal but is always much smaller than the correlation between the trading volume and the noise.

Hence, it seems that trading volume per se brings little information about the true value of the asset. As we shall see in the next session, this is valid only if we limit ourselves to extracting information from trading volume alone. This would mean ignoring another publicly available variable: the market price. The fact that the trading volume is correlated with the noise could be used to filter the noise from the price and use that to extract more information about the true value of the asset, as the next session will show.

4.4 Conditioning on price and volume

We compute $P(x > \lambda | p, z)$ using $\lambda = .2$ as in the previous section. The results of fixing one conditioning variable at a given level and increasing the other are displayed in Figure 4. In the upper panel, we fix p equal to $0.5\sigma_p$ (the solid line) and $-0.5\sigma_p$ (the dashed line) and vary z between 0.1 and 1.5. By way of comparison, we also graph the unconditional probability $P(x > .2)$ (the gray line). The conditional probability $p(x > .2|p, z)$ is increasing in z when $p = -0.5\sigma_p$ but decreasing in z when $p = 0.5\sigma_p$. This results also holds for other values of p and the effect on increasing z on $P(x > .2|p, z)$ when p is held fixed are always greater than on $p(x > .2|z)$

The middle panel of Figure 4 shows how $P(x > .2|p, z)$ varies when p increases for different levels of z: $z = E[z]$ (the solid line), $z = E[z] + 0.5\sigma_z$ (the long-dashed line), $z = E[z] - 0.5\sigma_z$ (the short-dashed line). Because of the symmetry of the conditional probability, only positive values of p are used. For average or below-average trading volume, the conditional probability conditioned on trading volume and price also increases with p. However, when trading volume is high, $P(x > .2|p, z)$ is non-monotonous in p: it can be a decreasing function of p for p below a threshold.

The conclusions drawn from graphing $P(x > .2|p, z)$ fixing one conditioning variable to some chosen level and letting the other vary are confirmed by looking at the contour plot of $P(x > .2|p, z)$ where both p and z vary. As z increases, $p(x > .2|p, z)$ rises for negative values of p and falls for positive values of p, while, for large levels of the trading volume, $P(x > .2|p, z)$ can be a decreasing function of p.

Figure 4 shows how the upper tail of the conditional distribution responds to changes in p and z. However, traders might be more interested in the conditional mean of x given p and z. Figure 5 displays how $E[x|p,z]$ reacts to changes in one conditioning variable when the other is held constant. The conclusions obtained about the tail of the conditional probability distribution also hold when looking at the conditional mean. $E[x|p,z]$ is an increasing function of z when the price is negative but a decreasing function of z when the price is positive. Note that the price has mean zero by construction, so that a positive (resp., negative) price should be interpreted in practice as an above-average (resp., below-average) price. In the market, a strong volume associated with an above-average price should be a bearish signal for the asset.

When the volume is about average, $E[x|p,z]$ is close to $E[x|p]$ and increases with p. The positive response of $E[x|p,z]$ to an increase in p is even more pronounced for lower levels of trading volume. However, when volume is strong, $E[x|p,z]$ can be a decreasing function of p (when $p > 0$) even though the market price and the true asset value are positively correlated.

Figures 4 and 5 show that one should be prudent when using the results of univariate conditioning in a multivariate context. For example, since $P(x > .2|z)$ and $P(x > .2|p)$ are increasing functions of z and p, it may seem intuitive to think that $P(x > .2|p,z)$ would be an increasing function of p and z. If this were true, one would want to adjust upward the value of the asset when the market price is high relative to its mean and, at the same time, the trading volume is strong. However, precisely the reverse is true.

This shows that, even though trading volume yields little information on the true value of the asset when used alone, it can be very useful when used in combination with the market price. Adding trading volume to the conditioning information set does not just fine-tune the results of using price alone; in some cases, it can overturn the conclusions obtained using a univariate conditioning scheme.

Looking at the joint distribution of price and volume may help one to relate univariate and multivariate conditioning, Figure 6 displays a contour plot of the joined density of x (resp. p) and z where x (or p) varies between -1.3 and 1.3 and z varies between 0 and 2.5. The joint density of p and z is shown in the upper panel. Even though both densities are symmetric, they display very different shapes. For high levels of the trading volume, the joint density of p and z is a bimodal function of p. Consequently, the conditional density of p on z (shown in the lower panel) becomes bimodal when trading volume is high enough. When the trading volume is low, the conditional density of of p given z is concentrated around its mean (which is 0 here). In contrast, for higher levels of the trading volume, the price is more likely to be either very low or very high than to be around average. Looking at the joint density of p and z makes it clear that the covariance between $|p|$ and z must be positive. To relate the univariate and bivariate conditioning results, recall that

$$\begin{aligned} P(x > \lambda|z)(h) &= \int_{t=-\infty}^{+\infty} P(x > \lambda|p,z)(t,h) f_{p|z}(t,h)\,dt \\ &= \int_{t=-\infty}^{0} P(x > \lambda|p,z)(t,h)\, f_{p|z}(t,h)\,dt \\ &+ \int_{t=0}^{+\infty} P(x > \lambda|p,z)(t,h)\, f_{p|z}(t,h)\,dt \end{aligned} \tag{7}$$

where $f_{p|z}$ is the conditional density function of p given z. $P(x > \lambda|p,z)(t,h)$ is an increasing function of h when t is negative and a decreasing function of h when t is positive, and $f_{p|z}(t,h) = f_{p|z}(-t,h)$. Intuitively, the effect of increasing h on $\int_{t=-\infty}^{0} P(x > \lambda|p,z)(t,h)\, f_{p|z}(t,h)\, dt$ and on $\int_{t=0}^{+\infty} P(x > \lambda|p,z)(t,h)\, f_{p|z}(t,h)\, dt$ should cancel out, to render $P(x > \lambda|z)$ rather insensitive to changes of z.

Using $E[x|z,p]$ to compute $E[x|z]$ shows even more clearly how the effect on $E[x|p,z]$ of an increase in z when $p > 0$ and when $p \leq 0$ cancel out.

$$
\begin{aligned}
E[x|z](h) &= \int_{t=-\infty}^{+\infty} E[x|p,z](t,h) f_{p|z}(t,h)\, dt \\
&= \int_{t=-\infty}^{0} E[x|p,z](t,h)\, f_{p|z}(t,h)\, dt + \int_{t=0}^{+\infty} E[x|p,z](t,h)\, f_{p|z}(t,h)\, dt \\
&= \int_{t=0}^{+\infty} \big(E[x|p,z](t,h) - E[x|p,z](-t,h)\big) f_{p|z}(t,h)\, dt \\
&= 0
\end{aligned}
\tag{8}
$$

Because of the symmetry of $E[x|p,z]$ with respect to p, $E[x|z] = 0$ even though varying z has a large (positive or negative) effect on $E[x|p,z]$ for a fixed p (for $p \neq 0$).

5 Kyle model

Kyle (1985) introduces a dynamic model of insider trading with a single risk-neutral informed trader, a noise trader and a competitive risk-neutral market maker. To translate Kyle's model into our notations, let y_1 be the demand of the informed trader, $y_2 = \varepsilon$ be the demand of the noise trader, $y_3 = -(y_1 + y_2)$ be the demand of the market maker. As before, x is the true value of the asset (observed by the informed trader) and p is the market price. The state variables, x and ε, are normally and independently distributed with means zero and respective variances σ_x^2 and σ_ε^2. The market maker observes the aggregate order flow and set the price equal to the conditional mean of the true value of the asset conditioned on this variable: $p = E[x|y_1 + y_2]$. In turn, the informed trader takes the market price into account to decide how much to trade. Under these conditions, there is a unique linear Nash equilibrium.

$$
\begin{aligned}
y_1 &= \beta x \\
y_2 &= \varepsilon \\
p &= E[x|y_1 + y_2] = \lambda(y_1 + y_2)
\end{aligned}
\tag{9}
$$

with $\beta = \sigma_\varepsilon/\sigma_x$ and $\lambda = 1/(2\beta)$. Kyle (1985) extends this static setup to dynamic games. However, we use only the static version of the model described by (9) because we want to compare $E[x|p,z]$ and $E[x|z]$ with results obtained in the static setup used in the previous section.

Kyle's framework has been extended and used by many financial economists. In these models, traders' demands, asset value and market prices are jointly normally distributed. The mean demands are zero, and the asset value mean (and hence the price mean) can be normalized to zero. Clearing the market by taking the market maker's net demand into account, Propositions 1 to 4 apply. Consequently, with v representing the level or the first difference of the price or the value of the asset, $E[v|z] = cov(v,z) = 0$, $cov(|v|;z) > 0$ and $p(v > \lambda|z)$ is increasing in z (the last result being based on numerical examples).

Proposition 5. *In the Kyle model, using price and volume does not yield more information about the conditional mean of the stock than using only the price*

$$E[x|z,p] \quad = \quad E[x|p] \quad = \quad p \tag{10}$$

However, trading volume can yield information on the tails of the distribution.

$$\begin{array}{lcll} P[x > s|z,p](z,p) & = & \frac{1}{2} I[z > s] & \text{for} \quad p > 0 \\ P[x > s|z,p](z,p) & = & \frac{1}{2} I[z > s + 2\beta|p|] & \text{for} \quad p \leq 0 \end{array} \tag{11}$$

The fact that $E[x|z,p] = E[x|p]$ may be due to the fact that the market price already equals the conditional value of x on the order flow.

6 Conclusion and further research

The contribution of this paper is to show how to use both the equilibrium price and trading volume to extract information about the true value of a traded asset. In most market microstructures models, the true value of the asset and the market price are normally distributed, so that conditioning on the price alone is easy. In contrast, conditioning on the trading volume is harder because it is a sum of truncated random variables[3]. However, when the traders demands are normally distributed, it is easy to compute the conditional distribution of the asset value conditioned on trading volume, or on trading volume and price. Once the signs of all the traders' demands are known, the trading volume becomes a sum of normally distributed random variables. Hence, the joint distribution of the trading volume and a given set of normally distributed random variables can be easily derived by considering every possible configuration for the signs of the traders' demands. Once the joint distribution of the trading volume, the market price, and the true asset value has been computed, it is easy to derive the conditional distribution of the asset value on volume and price.

We implement the method on a simple microstructures model. The main contribution of the microstructures model used is that the noise trader is price sensitive, which seems more realistic than modeling it as a pure random shock. Conditioning on both trading volume and price yields very different results than conditioning on those variables individually. For example, whereas the conditional mean of the true asset value on the market price is a linear, increasing function of price, the conditional expectation of the asset value on the market price and volume can be decreasing in price for a given value of the trading volume. Also, while the conditional mean of the asset value on trading volume coincides with the unconditional mean, the conditional mean conditioned on price and trading volume is decreasing in the trading volume when the market price is above average. Intuitively, in an efficient market, the equilibrium price should be more responsive to orders stemming from the informed traders than to those stemming from the noise trader.[4] This means that when the informed trader wants to trade on a positive realization of his private signal, his demand should

[3]The trading volume is the sum of all the the buy (resp. sell) orders

[4]Naturally, the trader's type is not revealed to the others. However, in equilibrium, the price is a (linear) functional of the private signals and the noise and reacts differently to the informed traders and the noise traders.

move the price against him, thus reducing the total amount of trade. In contrast, when the noise trader places buy orders, he should affect much less the market price, so that more of his potential demand can be satisfied. It follows that, for a given, above-average level of the market price, the higher is trading volume, the more the price should should reflect the buy orders of the noise traders. Hence, the observer interested in inferring the true value of the asset should use the trading volume as a proxy for noise trading and apply a discount when the volume is high.

Appendix

Proof of Proposition 1

$$
\begin{array}{rcl}
E[z\,x] & = & \sum_{i=1}^{n} E[x\,y_i^+] \\
& = & \sum_{i=1}^{n} E[x\,y_i\,I[y_i > 0]] \\
& = & \sum_{i=1}^{n} E[(-x)\,(-y_i)\,I[-y_i > 0]] \\
& = & \sum_{i=1}^{n} E[x\,y_i\,I[y_i \leq 0]] \\
& = & \sum_{i=1}^{n} E[x\,y_i^-] \\
& = & -E[z\,x]
\end{array}
\tag{12}
$$

The third line comes from the distribution of $(x, y_1, \ldots, y_n)$ being symmetric around zero; the fourth line from $P(y_i = 0) = 0$ all $i = 1, \ldots, n$. Hence $E[z\,x] = 0$ and $cov(x, z) = 0$.

Lemma 1: *Let* $u \sim N(0, 1)$

$$
\begin{array}{lcl}
E[u^+] & = & \frac{1}{\sqrt{2\pi}} \\
E[(u^+)^2] & = & \frac{1}{2} \\
\sigma(|u|) & = & \sqrt{1 - \frac{2}{\pi}}
\end{array}
\tag{13}
$$

If v_1 *and* v_2 *are jointly normally distributed with mean zero, variances* σ_1^2 *and* σ_2^2 *and correlation coefficient* ρ_{12}*,*

$$
\left\{
\begin{array}{lcl}
E[I[v_1 > 0, v_2 > 0]v_1^2] & = & \sigma_1^2 \left(\frac{1}{4} + \frac{1}{2\pi} \arcsin(\rho_{12}) + \frac{1}{2\pi} \rho_{12} \sqrt{1 - \rho_{12}^2} \right) \\
E[I[v_1 > 0, v_2 > 0]v_1 v_2] & = & \sigma_1 \sigma_2 \left(\rho_{12} (\frac{1}{4} + \frac{1}{2\pi} \arcsin(\rho_{12})) + \frac{1}{2\pi} \sqrt{1 - \rho_{12}^2} \right)
\end{array}
\right.
\tag{14}
$$

Proof To prove the first line of equation (13), recall that $E[u^+] = \int_0^\infty u\phi(u)du$ where ϕ is the standard normal density function. Since $\phi'(u) = -u\phi(u)$, $\int_0^\infty u\phi(u)du = [-\phi(u)]_0^\infty = \phi(0) = \frac{1}{\sqrt{2\pi}}$ To prove the second line, recall that $E[(u^+)^2] = E[I[u > 0]u^2]$. Using $E[I[u > 0]u^2] = E[I[u \leq 0]u^2]$ and $E[u^2] = 1$, we get the results. To obtain the third line, use the fact that $|u| = u^+ - u^-$, the symmetry of u^+ and u^- and $E[u^+ u^-] = 0$, to get $E[|u|] = 2E[u^+] = 1$ and $E[|u|^2] = 2E[(u^+)^2] = \sqrt{\frac{2}{\pi}}$.
To prove equation (14), let $G^1(v_1, v_2) = v_1 f(v_1, v_2)$, where f is the density function of v. Let $G_i^1 = \frac{\partial G^1}{\partial v_i}$, $i = 1, 2$. Then $G_1^1 = v_1 f_1 + f$, and $G_2^1 = v_1 f_2$. Hence,

$$
\left\{
\begin{array}{lclcl}
(\sigma^{11} v_1^2 + \sigma^{12} v_1 v_2) f & = & -v_1 f_1 & = & f - G_1^1 \\
(\sigma^{12} v_1^2 + \sigma^{22} v_1 v_2) f & = & -v_1 f_2 & = & -G_2^1
\end{array}
\right.
\tag{15}
$$

We will write $v = (v_1, v_2)$, $I = I[v_1 > 0, v_2 > 0]$, and $p(v > 0) = p(v_1 > 0, v_2 > 0)$. Noting that, for any value of v_2, $\int_{v_1=0}^\infty G_1^1(v_1, v_2)dv_1 = [G^1(v_1, v_2)]_{v_1=0}^\infty = 0$, we get

$$
\sigma^{11}\,E[I\,v_1^2] + \sigma^{12}\,E[I\,v_1 v_2] = p(v > 0). \tag{16}
$$

Also,

$$\begin{array}{rcl}\sigma^{12}\,E[I\,v_1^2]+\sigma^{22}\,E[I\,v_1v_2] & = & -\int_{v_1=0}^{\infty}\int_{v_2=0}^{\infty}\frac{\partial}{\partial v_2}G^1(v_1,v_2)dv_1dv_2 \\ & = & -\int_{v_1=0}^{\infty}[G^1(v_1,v_2)]_{v_2=0}^{\infty}dv_1 \\ & = & \frac{1}{2\pi\sqrt{|\Sigma|}}\int_{v_1=0}^{\infty}v_1e^{-\frac{1}{2}\sigma^{11}v_1^2}dv_1 \\ & = & \frac{1}{2\pi\sigma^{11}\sqrt{|\Sigma|}}\int_{u=0}^{\infty}ue^{-\frac{1}{2}u^2}du \\ & = & \frac{1}{2\pi\sigma^{11}\sqrt{|\Sigma|}} \\ & = & \frac{\sqrt{|\Sigma|}}{2\pi\sigma_{22}}\end{array} \tag{17}$$

Note that $\frac{1}{\sigma^{11}\sqrt{|\Sigma|}}=\frac{\sqrt{|\Sigma|}}{\sigma_{22}}$, as, by definition $\sigma^{11}=\frac{\sigma_{22}}{|\Sigma|}$. Hence,

$$\begin{pmatrix} E[I\,v_1^2] & E[I\,v_1v_2] \\ E[I\,v_1v_2] & E[I\,v_2^2]\end{pmatrix} = \Sigma\begin{pmatrix} p(v>0) & \frac{\sqrt{|\Sigma|}}{2\pi\sigma_{11}} \\ \frac{\sqrt{|\Sigma|}}{2\pi\sigma_{22}} & p(v>0)\end{pmatrix} \tag{18}$$

We can simplify notations, using $|\Sigma|=\sigma_1^2\sigma_2^2-\sigma_{12}^2$, which yields $\frac{\sqrt{|\Sigma|}}{\sigma_{22}}=\frac{\sigma_1}{\sigma_2}\sqrt{1-\rho_{12}^2}$. Equation (14) follows.

Proof of proposition 2: The formulas for $E[z]$ and $var(z)$ can be derived from lemma (1). The formula for $cov(|u|, z0$ is proven below. $cov(|u|,z)=2\,cov(u^+;z)$, and $z=\Sigma_{i=1}^6 I[v_2^i>0,v_3^i>0](v_2^i+v_3^i)$, where v_2^i, v_3^i replace v_1^i and v_2^i defined in table (1). Noting $(\tilde{v}_2^i,\tilde{v}_3^i)=(v_2^{7-i},v_3^{7-i})=-(v_2^i;v_3^i)$, $i=1,2,3$, one gets

$$z=\Sigma_{i=1}^3(I[v_2^i>0,v_3^i>0](v_2^i+v_3^i)+I[\tilde{v}_2^i>0;\tilde{v}_3^i>0](\tilde{v}_2^i+\tilde{v}_3^i)) \tag{19}$$

Hence, with $v_1=u$, one need only compute $cov(I[v_1>0]v_1,I[v_2>0,v_3>0](v_2+v_3))=E[I[v_1>0,v_2>0,v_3>0]v_1(v_2+v_3)]-E[I[v_1>0]v_1]E[I[v_2>0,v_3>0](v_2+v_3)]$, where $v=(v_1,v_2,v_3)$ is normally distributed with mean zero. If (y_1,y_2) is a normally distributed vector with mean zero and variance S, noting $s_i=\sqrt{s_{ii}}$, $E[I[y_1>0]y_1]=\frac{s_1}{\sqrt{2\pi}}$ and $E[I[y_1>0,y_2>0]y_1]=\varphi(S)$, with $\varphi(S)=\frac{1}{2\sqrt{2\pi}}(s_1+\frac{s_{12}}{s_2}$. Let $G(v_1,v_2,v_3)=v_1f(v_1,v_2;v_3)$ where f is the density function of v, $G_1(v)=v_1f_1(v)+f(v)$, then $G_2(v)=v_1f_2$, $G_3=v_1f_3$. In the following, $\Sigma=(\sigma_{ij})_{i;j=1}^3$,$\Sigma^{-1}=(\sigma^{ij})_{i,j=1}^3$, $\sigma_i=\sqrt{\sigma_{ii}}$, $\sigma^i=\sqrt{\sigma^{ii}}$. Γ_i^* is the matrix obtained by deleting the ith row and the ith column of Σ^{-1}, $v=(v_1,v_2,v_3)$. The first derivative of f with respect to v is $-(\Sigma^{-1}.v)f(v)$, hence

$$\left\{\begin{array}{rcl}(\sigma^{11}v_1^2+\sigma^{12}v_1v_2+\sigma^{13}v_1v_3)f(v) & = & -G_1(v)+f(v) \\ (\sigma^{21}v_1^2+\sigma^{22}v_1v_2+\sigma^{23}v_1v_3)f(v) & = & -G_2(v) \\ (\sigma^{31}v_1^2+\sigma^{32}v_1v_2+\sigma^{33}v_1v_3)f(v) & = & -G_3(v)\end{array}\right. \tag{20}$$

Now, take expectation through the system (27), beginning with the first row.

$$\int_{t>0}G_1(t)dt=\int_{t_2>0,t_3>0}[G(t)]_{t_1=0}^{+\infty}dt_2\,dt_3=0 \tag{21}$$

and

$$\int_{v>0}f(v)dv=p(v>0)=\frac{1}{8}+\frac{1}{4\pi}(\arcsin(\rho_{12})+\arcsin(\rho_{13})+\arcsin(\rho_{23})) \tag{22}$$

Taking expectation through the second row, one gets equation (30).

$$\begin{array}{rcl}\int_{t>0} -G_2 dt & = & \int_{t_1>0}\int_{t_3>0} -t_1[f(t)]_{t_2=0}^{+\infty} dt_1\, dt_3 \\ & = & \int_{t_1>0}\int_{t_3>0} t_1 \frac{1}{(2\pi)^{\frac{3}{2}}}\frac{1}{|\Sigma^{\frac{1}{2}}} \exp(-\frac{1}{2}(\sigma^{11}t_1^2 + 2\sigma^{13}t_1t_3 + \sigma^{33}t_3^2))dt_1\, dt_2 \\ & = & (2\pi|\Gamma_2^*||\Sigma|)^{-\frac{1}{2}} \int_{t_1>0}\int_{t_3>0} t_1 \frac{1}{(2\pi)}|\Gamma_2^*|^{\frac{1}{2}} \exp(-\frac{1}{2}(t_1,t_3)\Gamma_2^*(t_1,t_3)')dt_1 dt_3 \\ & = & (2\pi|\Gamma_2^*||\Sigma|)^{-\frac{1}{2}} E[I[\eta_1 > 0, \eta_2 > 0]\eta_1] \\ & = & \frac{1}{\sigma_2\sqrt{2\pi}} E[I[\eta_1 > 0, \eta_2 > 0]\eta_1] \\ & = & \frac{1}{\sigma_2\sqrt{2\pi}} \varphi(\Gamma_2^{*-1})\end{array} \tag{23}$$

where (η_1, η_2) is normally distributed with mean zero and variance Γ_2^{*-1}. The last but one line of (30) follows from the identity $\sigma_{22} = \frac{|\Gamma_2^*|}{\Sigma^{-1}}$, the last line from $E[I[\eta_1 > 0, \eta_2 > 0]\eta_1] = \varphi(\Gamma_2^{*-1})$. Likewise, $\int_{t>0} -G_3 dt = \frac{1}{\sigma_3\sqrt{2\pi}}\varphi(\Gamma_3^{*-1})$. Let $\rho_2^* = \frac{\sigma_{13}}{\sigma^1\sigma_3}$ and $\rho_3^* = \frac{\sigma_{12}}{\sigma^1\sigma_2}$. Using the definition of Γ_2^{*-1} and noting that $\frac{\sigma^{33}}{|\Gamma_2^*|} = \frac{1}{\sigma^{11}(1-(\rho_2^*)^2)}$, one gets $\varphi(\Gamma_2^{*-1}) = \frac{1}{2\sigma^1\sqrt{2\pi}}\sqrt{\frac{1-\rho_2^*}{1+\rho_2^*}}$. Finally

$$\begin{pmatrix} E[I[v>0]v_1^2] \\ E[I[v>0]v_1v_2] \\ E[I[v>0]v_1v_3] \end{pmatrix} = \Sigma \begin{pmatrix} \frac{1}{8} + \frac{1}{4\pi}(\arcsin(\rho_{12}) + \arcsin(\rho_{13}) + \arcsin(\rho_{23})) \\ \frac{1}{4\pi\sigma_2\sigma^1}\sqrt{\frac{1-\sigma_2^*}{1+\sigma_2^*}} \\ \frac{1}{4\pi\sigma_3\sigma^1}\sqrt{\frac{1-\sigma_3^*}{1+\sigma_3^*}} \end{pmatrix} \tag{24}$$

and

$$\begin{array}{rcl} E[I[v>0]v_1(v_2+v_3)] & = & (\sigma_{12}+\sigma_{13})(\frac{1}{8} + \frac{1}{4\pi}(\arcsin(\rho_{12}) + \arcsin(\rho_{13}) + \arcsin(\rho_{23}))) \\ & + & \frac{1}{4\pi\sigma^1}\sqrt{\frac{1-\sigma_2^*}{1+\sigma_2^*}}(\sigma_2 + \frac{\sigma_{23}}{\sigma_2}) + \frac{1}{4\pi\sigma^1}\sqrt{\frac{1-\sigma_3^*}{1+\sigma_3^*}}(\sigma_3 + \frac{\sigma_{23}}{\sigma_3}) \end{array} \tag{25}$$

Let $\tilde{v} = (\tilde{v}_1, \tilde{v}_2, \tilde{v}_3) = (v_1, -v_2, -v_3)$ and $var(\tilde{v}) = \tilde{\Sigma} = (\tilde{\sigma}_{ij})_{i,j=1}^3$. Naturally, $\tilde{\rho}_{12} = -\rho_{12}$, $\tilde{\rho}_{13} = -\rho_{13}$, $\tilde{\rho}_{23} = \rho_{23}$. Writing $\tilde{\Sigma}^{-1} = (\tilde{\sigma}_{ij})_{i,j=1}^3$ and using the definition of the inverse, one also gets $\tilde{\sigma}^{ii} = \sigma^{ii}$ for $i = 1,2,3$, $\tilde{\sigma}^{12} = -\sigma^{12}$, $\tilde{\sigma}^{13} = -\sigma^{13}$, and $\tilde{\sigma}^{23} = \sigma^{23}$, and consequently $\tilde{\rho}_2^* = -\rho_2^*$, $\tilde{\rho}_3^* = -\rho_3^*$. Hence

$$\begin{array}{rcl} E[I[\tilde{v}>0]\tilde{v}_1(\tilde{v}_2+\tilde{v}_3)] & = & -(\sigma_{12}+\sigma_{13})(\frac{1}{8} + \frac{1}{4\pi}(-\arcsin(\rho_{12}) - \arcsin(\rho_{13}) + \arcsin(\rho_{23}))) \\ & + & \frac{1}{4\pi\sigma^1}\sqrt{\frac{1+\sigma_2^*}{1-\sigma_2^*}}(\sigma_2 + \frac{\sigma_{23}}{\sigma_2}) + \frac{1}{4\pi\sigma^1}\sqrt{\frac{1+\sigma_3^*}{1-\sigma_3^*}}(\sigma_3 + \frac{\sigma_{23}}{\sigma_3}) \end{array} \tag{26}$$

and

$$\begin{array}{l} E[I[v>0]v_1(v_2+v_3)] + E[I[\tilde{v}>0]\tilde{v}_1(\tilde{v}_2+\tilde{v}_3)] = \\ (\sigma_{12}+\sigma_{13})\frac{1}{2\pi}(\arcsin(\rho_{12}) + \arcsin(\rho_{13})) + \\ \frac{1}{4\pi\sigma^1}(\sigma_2 + \frac{\sigma_{23}}{\sigma_2})\left(\sqrt{\frac{1-\rho_2^*}{1+\rho_2^*}} + \sqrt{\frac{1+\rho_2^*}{1-\rho_2^*}}\right) + \frac{1}{4\pi\sigma^1}(\sigma_2 + \frac{\sigma_{23}}{\sigma_2})\left(\sqrt{\frac{1-\rho_3^*}{1+\rho_3^*}} + \sqrt{\frac{1+\rho_3^*}{1-\rho_3^*}}\right) \end{array} \tag{27}$$

$$\begin{array}{rcl} \frac{1}{2\sigma^1}\left(\sqrt{\frac{1-\rho_2^*}{1+\rho_2^*}} + \sqrt{\frac{1+\rho_2^*}{1-\rho_2^*}}\right) & = & \frac{1}{\sigma^1\sqrt{1-(\rho_2^*)^2}} \\ & = & \sqrt{\frac{\sigma^{33}}{|\Gamma_2^*|}} \\ & = & \sqrt{\frac{\sigma^{33}|\Sigma|}{\sigma_{22}}} \\ & = & \sqrt{\frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{22}}} \\ & = & \sigma_1\sqrt{1-\rho_{12}^2} \end{array} \tag{28}$$

and

$$\begin{array}{l} E[I[v>0]v_1(v_2+v_3)]+E[I[\tilde{v}>0]\tilde{v}_1(\tilde{v}_2+\tilde{v}_3)]= \\ (\sigma_{12}+\sigma_{13})(\frac{1}{2\pi}(\arcsin(\rho_{12})+\arcsin(\rho_{13}))+ \\ \frac{1}{2\pi}\sigma^1(\sigma_2+\frac{\sigma_{23}}{\sigma_2}\sqrt{1-\rho_{12}^2})+\frac{1}{2\pi}\sigma^1(\sigma_3+\frac{\sigma_{23}}{\sigma_3}\sqrt{1-\rho_{13}^2}) \end{array} \tag{29}$$

Since

$$E[I[v_1>0]v_1]=\frac{1}{\sqrt{2\pi}}\sigma_1 \tag{30}$$

and

$$E[I[v_2>0,v_3>0](v_2+v_3)]=E[I[\tilde{v}_2>0;\tilde{v}_3>0](\tilde{v}_2+\tilde{v}_3)]=\frac{1}{2\sqrt{2\pi}}\left(\sigma_2+\frac{\sigma_{23}}{\sigma_3}+\sigma_3+\frac{\sigma_{23}}{\sigma_2}\right) \tag{31}$$

$$cov(I[v_1>0]v_1,I[v_2>0,v_3>0](v_2+v_3))+cov(I[\tilde{v}_1>0]\tilde{v}_1,I[\tilde{v}_2>0,\tilde{v}_3>0](\tilde{v}_2+\tilde{v}_3))= \frac{\sigma_1}{2\pi}(H(\Sigma)+G(\Sigma)) \tag{32}$$

$$\begin{array}{rcl} H(\Sigma) & = & \left(\frac{\sigma_{12}}{\sigma_1}+\frac{\sigma_{13}}{\sigma_1}\right)(\arcsin(\rho_{12})+\arcsin(\rho_{13})) \\ G(\Sigma) & = & \left(\sigma_2+\frac{\sigma_{23}}{\sigma_2}\right)\left(\sqrt{1-\rho_{12}^2}-1\right)+\left(\sigma_3+\frac{\sigma_{23}}{\sigma_3}\right)\left(\sqrt{1-\rho_{13}^2}-1\right) \end{array} \tag{33}$$

Let $v_1=u$, and (v_2,v_3) can take three values; in the first case $(v_2,v_3)=(y_1,y_2)$, in the second case, $(v_2,v_3)=(-y_2,y_1+y_2)$, in the third case $(v_2,v_3)=(y_1,-(y_1+y_2))$. Call Σ_i the variance matrix of v in case i, $H=H(\Sigma_1)+H(\Sigma_2)+H(\Sigma_3)$, and $G=G(\Sigma_1)+G(\Sigma_2)+G(\Sigma_3)$. Then $cov(u^+,z)=\frac{\sigma_u}{2\pi}(H+G)$.

$$H=\sigma_u^{-1}\left(cov(u,y_1)\arcsin(\rho_{u,y_1})+cov(u,y_2)\arcsin(\rho_{u,y_2})+cov(u,y_1+y_2)\arcsin(\rho_{u,y_1+y_2})\right) \tag{34}$$

This is because

$$\begin{array}{l} \sigma_u H=cov(u,y_1+y_2)[\arcsin(\sigma_{u,y_1})+\arcsin(\sigma_{u,y_2})] \\ +cov(u,y_1)[-\arcsin(\sigma_{u,y_2})+\arcsin(\sigma_{u,(y_1+y_2)})] \\ -cov(u;y_2)[-\arcsin(\sigma_{u,(y_1+y_2)})+\arcsin(\sigma_{u,y_1})] \end{array} \tag{35}$$

which rearranging terms, equals

$$cov(u,y_1)\arcsin(\sigma_u,y_1)+cov(u,y_2)\arcsin(\sigma_u,y_2)+cov(u,y_1+y_2)\arcsin(\sigma_u,.y_1+y2_) \tag{36}$$

Likewise,

$$G=\sigma_{y_1}(\sqrt{1-\rho_{u,y_1}^2}-1)+\sigma_{y_2}(\sqrt{1-\rho_{u,y_2}^2}-1)+\sigma_{y_1+y_2}(\sqrt{1-\rho_{u,y_1+y_2}^2}-1) \tag{37}$$

Finally,

$$cov(|u|,z)=\frac{\sigma_u}{\pi}(\sigma_{y_1}h(\rho_{u,y_1})+\sigma_{y_2}h(\rho_{u,y_2})+\sigma_{y_1+y_2}h(\rho_{u,y_1+y_2})) \tag{38}$$

with $h(t)=t\arcsin(t)+\sqrt{1-t^2}-1$. For all $t\in[-1,1]$, $h(t)\geq 0$ and $h(t)=0$ if and only if $t=0$. Consequently, $cov(|u|,z)$ is non-negative and is zero only if u is independent of (y_1,y_2)

Proof of Proposition 3. Let's show that

$$f_{p,z}(t,h) = \sum_{i=1}^{m} \int_{s=0}^{h} f^i(t,s,s-h)ds \tag{39}$$

the proof for $f_{x,p,z}$ is similar. We know that

$$P(|z-h|<\varepsilon, |p-t|<\eta) = \sum_{i=1}^{m} P(y\in A_i, |z-h|<\varepsilon, |p-t|<\eta) \tag{40}$$

Take the case $i=1$, let $f^1 = f$ be the density of (y_1, y_2, p) and show that

$$\frac{1}{2\varepsilon}\frac{1}{2\eta}P(y\in A_1, |z-h|<\varepsilon, |p-t|<\eta) \to \int_{y_1=0}^{h} f^1(y_1, h-y_1, t)dy_1 \tag{41}$$

If the equation above holds, the same will be true for $i>1$ and equation (39) and proposition (3) will be proven.

$$\begin{aligned} p(y\in A_1, |z-h|<\varepsilon, |p-t|<\eta) &= p(y_1>0, y_2>0, |z-h|<\varepsilon, |p-t|<\eta) \\ &= p(0<y_1<h-\varepsilon, |y_2-(h-y_1)|<\varepsilon, |p-t|<\eta) \\ &+ p(h-\varepsilon<y_1<h+\varepsilon, 0<y_2<h+\varepsilon-y_1, |p-t|<\eta) \end{aligned} \tag{42}$$

As ε and η converge to 0, the term in the first line converges to $f(y_1, h-y_1, t)$. and the term in the second line to 0.

$$\begin{aligned} &\tfrac{1}{2\varepsilon}\tfrac{1}{2\eta}p(0<y_1<h-\varepsilon, |y_2-(h-y_1)|<\varepsilon, |p-t|<\eta) \\ &= \int_{y_1=0}^{h-\varepsilon}\big[\tfrac{1}{2\varepsilon}\int_{y_2=h-y_1-\varepsilon}^{y_2=h-y_1+\varepsilon}\big(\tfrac{1}{2\eta}\int_{p=t-\eta}^{p=t+\eta} f(y_1,y_2,p)dp\big)dy_2\big]dy_1 \end{aligned} \tag{43}$$

As $\eta\to 0$, the term within the inner brackets converges to $f(y_1, y_2, t)$, substituting this expression in the equation and letting now $\varepsilon \to 0$, we get

$$\frac{1}{2\varepsilon}\frac{1}{2\eta}p(0<y_1<h-\varepsilon, |y_2-(h-y_1)|<\varepsilon, |p-t|<\eta) \to \int_{y_1=0}^{h} f(y_1, h-y_1, t)dy_1 \tag{44}$$

The remaining term converges to 0, as is shown below.

$$\begin{aligned} &\tfrac{1}{2\varepsilon}\tfrac{1}{2\eta}p(h-\varepsilon<y_1<h+\varepsilon, 0<y_2<h+\varepsilon-y_1, |p-t|<\eta) \\ &= \int_{y_1=h-\varepsilon}^{h+\varepsilon}\big[\tfrac{1}{2\varepsilon}\big(\int_{y_2=0}^{y_2=h-y_1+\varepsilon}\tfrac{1}{2\eta}\int_{p=t-\eta}^{p=t+\eta} f(y_1,y_2,p)dp\big)dy_2\big]dy_1 \\ &\le \int_{y_1=h-\varepsilon}^{h+\varepsilon}\big[\tfrac{1}{2\varepsilon}\big(\int_{y_2=0}^{y_2=2\varepsilon}\tfrac{1}{2\eta}\int_{p=t-\eta}^{p=t+\eta} f(y_1,y_2,p)dp\big)dy_2\big]dy_1 \\ &\le \int_{y_1=h-\varepsilon}^{h+\varepsilon}\big[\tfrac{1}{2\eta}\int_{p=t-\eta}^{p=t+\eta} f(y_1,\theta(\varepsilon),p)dp\big]dy_1 \end{aligned} \tag{45}$$

The 3rd line stems from $0<h-y_1<2\varepsilon$, the 4th line stems from the mean-value theorem. The term in the inner bracket of the last equation converges to $f(y_1, \theta(\varepsilon), t)$ as $\eta\to 0$, which is bounded, hence when $\varepsilon\to 0$, the integral converges to 0. We conclude that, when $\varepsilon\to 0$ and $\eta \to 0$, $\frac{1}{2\varepsilon}\frac{1}{2\eta}p(h-\varepsilon<y_1<h+\varepsilon, 0<y_2<h+\varepsilon-y_1, |p-t|<\eta)\to 0$.

Proof of Proposition 3. First, we show that

$$p(z\in H, p\in T) = p(z\in H, -p\in T) \tag{46}$$

The proof of $p(z \in H, p \in T, x \in U) = p(z \in H, -p \in T, -x \in U)$ is similar. The random variable z coincides with v_i when $y \in A_i$, $i = 1, \ldots, n$, (v_i, y, p) are symmetrically distributed around zero. Furthermore, $v_{n+1-i} = -v_i$ and $[y_i \in A_i] = [-y_{n+1-i} \in A_{n+1-i}]$.

$$\begin{array}{rcl} p(z \in H, p \in T) & = & \sum_{i=1}^{n} p(y \in A_i, z \in H, p \in T) \\ & = & \sum_{i=1}^{n} p(y \in A_i, v_i \in H, p \in T) \\ & = & \sum_{i=1}^{n} p(-y \in A_i, -v_i \in H, -p \in T) \\ & = & \sum_{i=1}^{n} p(y \in A_{n+1-i}, v_{n+1-i} \in H, -p \in T) \\ & = & \sum_{j=1}^{n} p(y \in A_j, v_j \in H, -p \in T) \\ & = & \sum_{j=1}^{n} p(y \in A_j, z \in H, -p \in T) \\ & = & p(z \in H, -p \in T) \end{array} \tag{47}$$

Hence,

$$\begin{array}{lcl} f_{x,p,z}(-u, -t, h) & = & f_{x,p,z}(u, t, h) \\ f_{x,z}(-u, h) & = & f_{x,z}(u, h) \\ f_{p,z}(-t, h) & = & f_{p,z}(t, h) \end{array} \tag{48}$$

Proposition (3)follows from equation (48) and the fact that

$$\begin{array}{lcl} E[x|p, z](t, h) & = & \int_{u=-\infty}^{+\infty} u\, f_{x|p,z}(u, t, h) du \\ P(x > \lambda|p, z)(t, h) & = & \int_{u=\lambda}^{+\infty} f_{x|p,z}(u, t, h) du \\ E[x|z](h) & = & \int_{u=-\infty}^{+\infty} u\, f_{x|z}(u, h) du \\ P(x > \lambda|z)(h) & = & \int_{u=\lambda}^{+\infty} f_{x|z}(u, h) du \end{array} \tag{49}$$

Proof of Proposition 5. The price is $p = \lambda(y_1 + y_2)$ and the trading volume is $z = \frac{1}{2}(|y_1| + |y_2| + |y_1 + y_2|)$, hence observing p and z is equivalent to observing $y_1 + y_2 = 2\,\beta p$ and $|y_1| + |y_2| = 2\,(z - \beta|p|)$ where $y_1 = \beta\, x$ is the informed trader's demand and y_2 is the noise. Knowing that y_1 and y_2 are of different sign, we can infer the value of y_1 (and hence that of x) from the realizations of z and p. In contrast, when y_1 and y_2 are of the same sign, observing z and p gives us information only on $y_1 + y_2$.

Let $a = y_1 + y_2$ and $b = |y_1| + |y_2|$. Since $|y_1| + |y_2| \geq |y_1 + y_2|$, $b \geq |a|$, i.e., $-b \leq a \leq b$. If $-b < a < b$, then y_1 and y_2 must be of different signs, although the sign of each variable cannot be determined. When $(y_1 > 0, y_2 \leq 0)$, $(y_1, y_2) = ((a+b)/2, (a-b)/2)$, when $(y_1 \leq 0, y_2 > 0)$, $(y_1, y_2) = ((a-b)/2, (a+b)/2))$,

Let $B(a, \eta)$ and $B(b, \varepsilon)$ be the open ball with centers a and b and radius η and ε.

$$\begin{array}{l} p(y_1 + y_2 \in B(a,\eta), |y_1| + |y_2| \in B(b,\varepsilon)) = \\ p(y_1 > 0, y_2 > 0, y_1 + y_2 \in B(a,\eta), |y_1| + |y_2| \in B(b,\varepsilon)) + \\ p(y_1 > 0, y_2 \leq 0, y_1 + y_2 \in B(a,\eta), |y_1| + |y_2| \in B(b,\varepsilon)) + \\ p(y_1 \leq 0, y_2 > 0, y_1 + y_2 \in B(a,\eta), |y_1| + |y_2| \in B(b,\varepsilon)) + \\ p(y_1 \leq 0, y_2 \leq 0, y_1 + y_2 \in B(a,\eta), |y_1| + |y_2| \in B(b,\varepsilon)) \end{array} \tag{50}$$

That is,

$$\begin{array}{l} p(y_1 + y_2 \in B(a,\eta), |y_1| + |y_2| \in B(b,\varepsilon)) = \\ p(y_1 > 0, y_2 > 0, y_1 + y_2 \in B(a,\eta) \cap B(b,\varepsilon)) + \\ p(y_1 > 0, y_2 \leq 0, y_1 + y_2 \in B(a,\eta), y_1 - y_2 \in B(b,\varepsilon)) + \\ p(y_1 \leq 0, y_2 > 0, y_1 + y_2 \in B(a,\eta), y_2 - y_1 \in B(b,\varepsilon)) + \\ p(y_1 \leq 0, y_2 \leq 0, y_1 + y_2 \in B(a,\eta) \cap B(-b,\varepsilon)) \end{array} \tag{51}$$

For $a \neq b$ and $a \neq -b$, we can find η and ε small enough so that $B(a,\eta) \cap B(b,\varepsilon)) = \emptyset$ and $B(a,\eta) \cap B(-b,\varepsilon)) = \emptyset$, so that

$$\begin{array}{l} p(y_1 + y_2 \in B(a,\eta), |y_1| + |y_2| \in B(b,\varepsilon)) = \\ p(y_1 > 0, y_2 \leq 0, y_1 + y_2 \in B(a,\eta), y_1 - y_2 \in B(b,\varepsilon)) + \\ p(y_1 \leq 0, y_2 > 0, y_1 + y_2 \in B(a,\eta), y_2 - y_1 \in B(b,\varepsilon)) = \\ p(y_1 > 0, y_2 \leq 0, y_1 \in B((a+b)/2, (\eta+\varepsilon)/2), y_2 \in B((a-b)/2, (\eta+\varepsilon)/2)) + \\ p(y_1 \leq 0, y_2 > 0, y_1 \in B((a-b)/2, (\eta+\varepsilon)/2), y_2 \in B((a+b)/2, (\eta+\varepsilon)/2)) = \\ p(y_1 \in B((a+b)/2, (\eta+\varepsilon)/2), y_2 \in B((a-b)/2, (\eta+\varepsilon)/2)) + \\ p(y_1 \in B((a-b)/2, (\eta+\varepsilon)/2), y_2 \in B((a+b)/2, (\eta+\varepsilon)/2)) = \end{array} \tag{52}$$

The last two lines stem from the fact that, since $a + b > 0 > a - b$, for η and ε small enough, $(a+b)/2 - (\eta+\varepsilon)/2 > 0$ and $(a-b)/2 - (\eta+\varepsilon)/2 < 0$. Hence, if $f(y_1, y_2)$ is the density function of (y_1, y_2), then, dividing both the numerator and the denominator by $\eta + \varepsilon$ and taking the limit, we get:

$$\begin{array}{l} \lim_{\eta,\varepsilon\to 0} \frac{P(y_1>0, y_2\leq 0, y_1+y_2\in B(a,\eta), y_1-y_2\in B(b,\varepsilon))}{p(y_1+y_2\in B(a,\eta), |y_1|+|y_2|\in B(b,\varepsilon))} = \frac{f((a+b)/2,(a-b)/2)}{f((a+b)/2,(a-b)/2)+f((a-b)/2,(a+b)/2)} \\ \lim_{\eta,\varepsilon\to 0} \frac{P(y_1\leq 0, y_2>0, y_1+y_2\in B(a,\eta), y_2-y_1\in B(b,\varepsilon))}{p(y_1+y_2\in B(a,\eta), |y_1|+|y_2|\in B(b,\varepsilon))} = \frac{f((a-b)/2,(a+b)/2)}{f((a+b)/2,(a-b)/2)+f((a-b)/2,(a+b)/2)} \end{array} \tag{53}$$

To summarize, for $-b < a < b$,

$$E[\phi(x)|y_1 + y_2, |y_1| + |y_2|](a,b) \;=\; \frac{f((a+b)/2,(a-b)/2)\,\phi((a+b)/2) + f((a-b)/2,(a+b)/2)\,\phi((a-b)/2)}{f((a+b)/2,(a-b)/2)+f((a-b)/2,(a+b)/2)} \tag{54}$$

Since $y_1 = \beta x$ with $\beta = \sigma_\varepsilon/\sigma_x$, y_1 and y_2 are jointly normally distributed random variables with the same mean and the same variance. Hence, they are identically distributed and the distribution of (y_1, y_2) is symmetric. Moreover,

$$\begin{array}{rclcl} a & = & y_1 + y_2 & = & 2\beta p \\ b & = & |y_1| + |y_2| & = & 2(z - \beta|p|) \end{array} \tag{55}$$

Since $\lambda = 1/(2\beta)$, $|a| < b$ is equivalent to $|p| < \lambda z$ and

$$\begin{array}{rcl} (a+b)/2 & = & z + (p - |p|)\beta \\ (a-b)/2 & = & -z + (p + |p|)\beta \end{array} \tag{56}$$

Hence,

$$E[\phi(x)|z,p](z,p) = \frac{1}{2}\left(\phi(z + (p-|p|)\beta) + \phi(-z + (p+|p|)\beta)\right) \tag{57}$$

Plugging in $\phi(x) = x$, we obtain $E[x|z,p] = p$. Moreover, $E[x|p] = p$ too. To see that note that because $p = \lambda(y_1 + y_2) = \lambda(\beta x + \varepsilon) = 1/2(x + \sigma_x/\sigma_\varepsilon \varepsilon)$, $E[xp]/E[p^2] = 1$. Hence

$$E[x|z,p] = E[x|p] = p \tag{58}$$

Let now $\phi(x) = I[x > s]$ with $s > 0$. Recall that $|p| \leq \lambda z$ and we assume that $|p| < \lambda z$. When $p > 0$, $z+(p-|p|)\beta = z > 0$ or $-z+(p+|p|)\beta = 2\beta(-\lambda z+|p|) < 0$, and hence $E[\phi(x)|z,p](z,p) = \frac{1}{2}\phi(z)$. When $p < 0$, $z+(p-|p|)\beta = 2\beta(\lambda z-|p|) > 0$ or $-z + (p + |p|)\beta = -z < 0$, and hence $E[\phi(x)|z,p](z,p) = \frac{1}{2}\phi(z - 2\beta|p|)$. Consequently,

$$\begin{array}{rcll} P[x > s||z,p](z,p) & = & \frac{1}{2}I[z > s] & \text{for} \quad p > 0 \\ P[x > s||z,p](z,p) & = & \frac{1}{2}I[z > s + 2\beta|p|] & \text{for} \quad p < 0 \end{array} \tag{59}$$

Proof of:

$$\left\{ \begin{array}{rcl} corr(|x|, z) & = & 0 \\ corr(|p|, z) & = & 1 \end{array} \right. \tag{60}$$

when $\rho = 0$.

In the models we work with, the traders' demands, $y_1, \ldots, y_{n+1}$ and the market price p are linear functions of the state variables $G_1, \ldots, G_n$ and ε. When there is no private information, the demands and the market price are functions of ε only: (in part 1 about the model, look at formula when $\rho \to 0$).

$$y_i = d_i\,\varepsilon\ ,\ i = 1, \ldots, n+1 \tag{61}$$

$$z = \sum_{i=1}^{n} y_i^+ = \sum_{i=1}^{n} (d_i\,\varepsilon)^+ \tag{62}$$

Let I and $\bar{I}$ be the sets so that

$$\left\{ \begin{array}{lll} d_i > 0 & \text{if} & i \in I \\ d_i < 0 & \text{if} & i \in \bar{I} \end{array} \right. \tag{63}$$

Note that we can just omit from the analysis the y_i with $d_i = 0$.

$$(d_i\,\varepsilon)^+ = \begin{cases} d_i\,\varepsilon^+ & \text{if} \quad d_i > 0 \\ d_i\,\varepsilon^- & \text{if} \quad d_i < 0 \end{cases} \tag{64}$$

Because the market clears, we have $\sum_{i=1}^{n+1} y_i = 0$, that is,

$$\sum_{i \in I} d_i = -\sum_{i \in \bar{I}} d_i \tag{65}$$

Using equations (64) and (65), we get

$$\begin{aligned} z &= \textstyle\sum_{i=1}^{n+1} y_i^+ \\ &= \textstyle\sum_{i=1}^{n+1} (d_i\,\varepsilon)^+ \\ &= \textstyle\sum_{i \in I} d_i\,\varepsilon^+ + \sum_{i \in \bar{I}} d_i\,\varepsilon^- \\ &= \textstyle\sum_{i \in I} d_i\,(\varepsilon^+ - \varepsilon^-) \\ &= d\,|\varepsilon| \end{aligned} \tag{66}$$

where $d = \sum_{i \in I} d_i \neq 0$. Since the market price is also a linear function of ε, we have $|p| = c|\varepsilon|$, with $c \neq 0$, which implies that $corr(z, |p|) = 1$. Naturally, since $E[x\,\varepsilon] = 0$ and $E[\varepsilon] = 0$, we have $corr(x, \varepsilon) = 0$.

cases (i)	sign of y_1	sign of y_2	sign of y_1+y_2	z	v_1^i	v_2^i
1	+	+	+	y_1+y_2	y_1	y_2
2	+	-	+	y_1	$-y_2$	y_1+y_2
3	+	-	-	$-y_2$	y_1	$-(y_1+y_2)$
4	-	+	+	y_2	$-y_1$	y_1+y_2
5	-	+	-	$-y_1$	y_2	$-(y_1+y_2)$
6	-	-	-	$-(y_1+y_2)$	$-y_1$	$-y_2$

Table 1: Decomposition of trading volume

There are three traders; y_1 is the first trader's demand, y_2 is the second trader's demand, the last trader's demand is $y_3 = -(y_1 + y_2)$. The trading volume is $z = y_1^+ + y_2^+ + y_3^+$. In each case i, $i = 1, \ldots, 6$, z can be written as $v_1^i + v_2^i$.

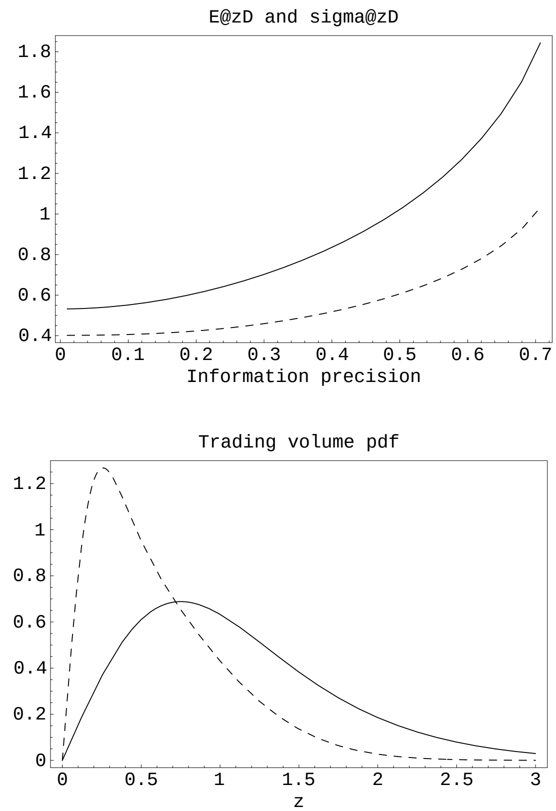

Figure 1: **Characteristics of the trading volume distribution.** Upper panel: $E[z]$ (solid line) and σ_z (dashed line) as information precision (ρ) increases. Lower panel: density of z for $\rho = .5$ (solid line) and $\rho = .2$ (dashed line).

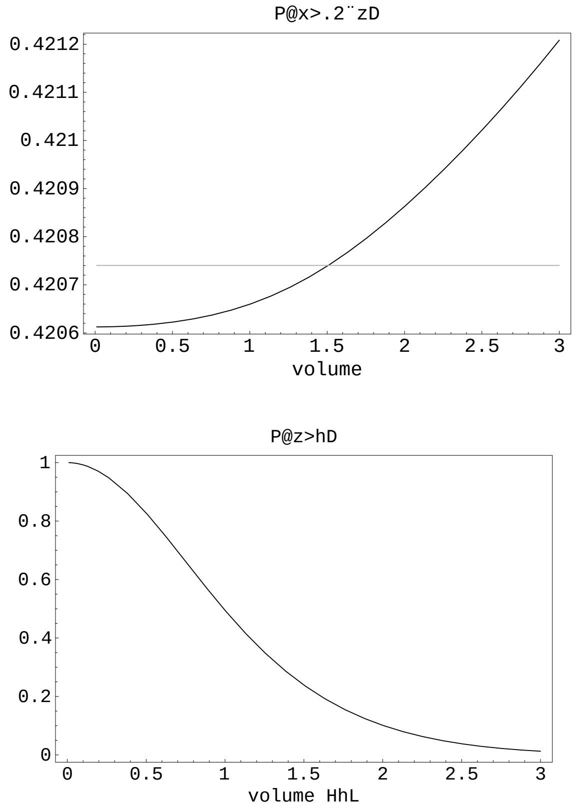

Figure 2: **Conditional tail probability conditioned on trading volume only.** Upper panel: $P(x > .2|z)$ (black line) as a function of z and $P(x > .2)$ (gray line). Lower panel: $(P(z > h)$ as a function of h.

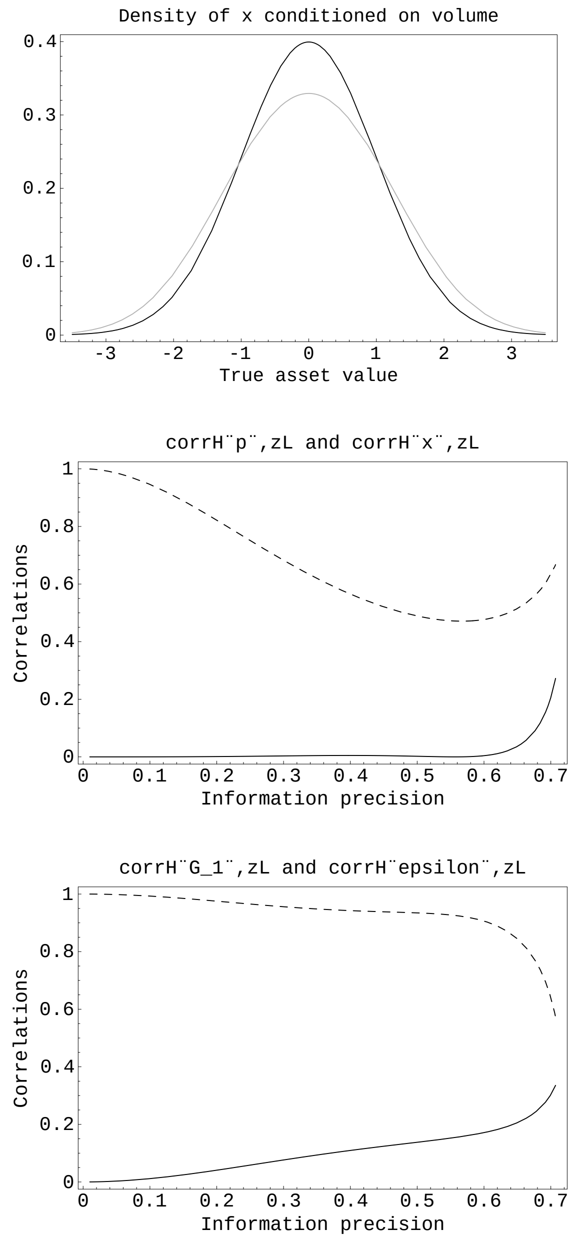

Figure 3: **Relations between trading volume, price and asset value.** Upper panel: conditional density of x given z for extremely high volume (gray line) and unconditional density function of x (black line). Middle panel: correlation between z and p (solid line) and between z and x (dashed line) as information precision ρ increases. Bottom panel: correlation between z and G_1 (solid line) and between z and ε (dashed line) as information precision ρ increases.

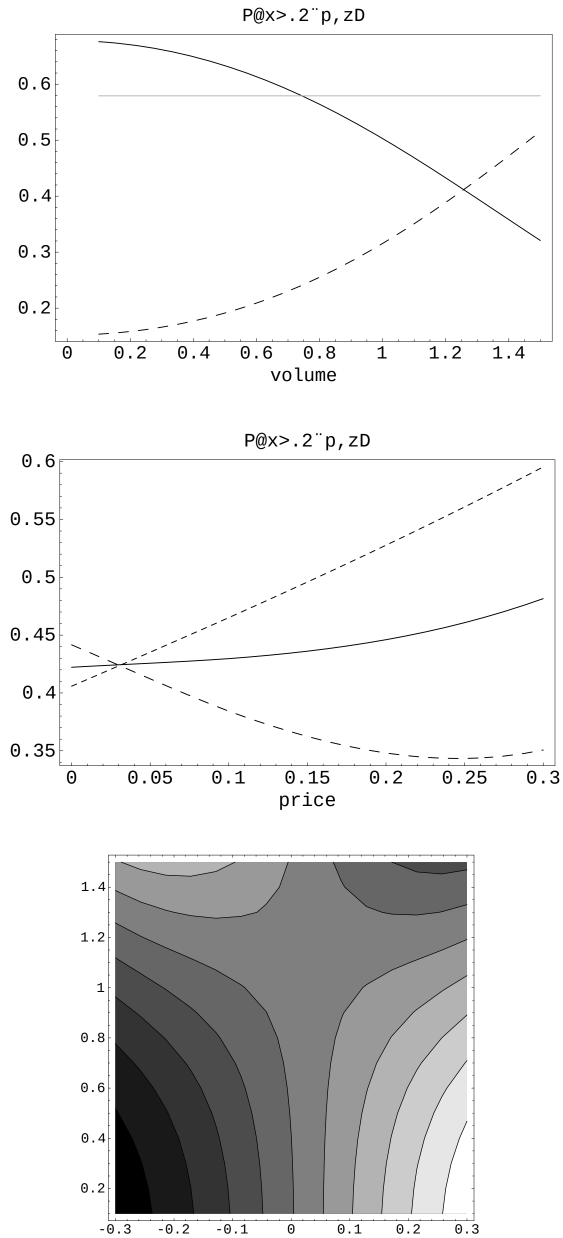

Figure 4: **Conditional tail probability conditioned on trading volume and price.** Upper panel: $P(x > .2|p, z)$ as a function of z when $z = 0.5\sigma_p$ (solid line) and $z = -0.5\sigma_p$ (dashed line) and $P(x > .2)$ (gray line). Middle panel: $P(x > .2|p, z)$ as a function of p increases for different levels of z: $z = E[z]$ (the solid line), $z = E[z] + 0.5\sigma_z$ (the dashed line), $z = E[z] - 0.5\sigma_z$ (the short-dashed line). Lower panel: $P(x > .2|p, z)$ as a function of p (horizontal axis) and z (vertical axis).

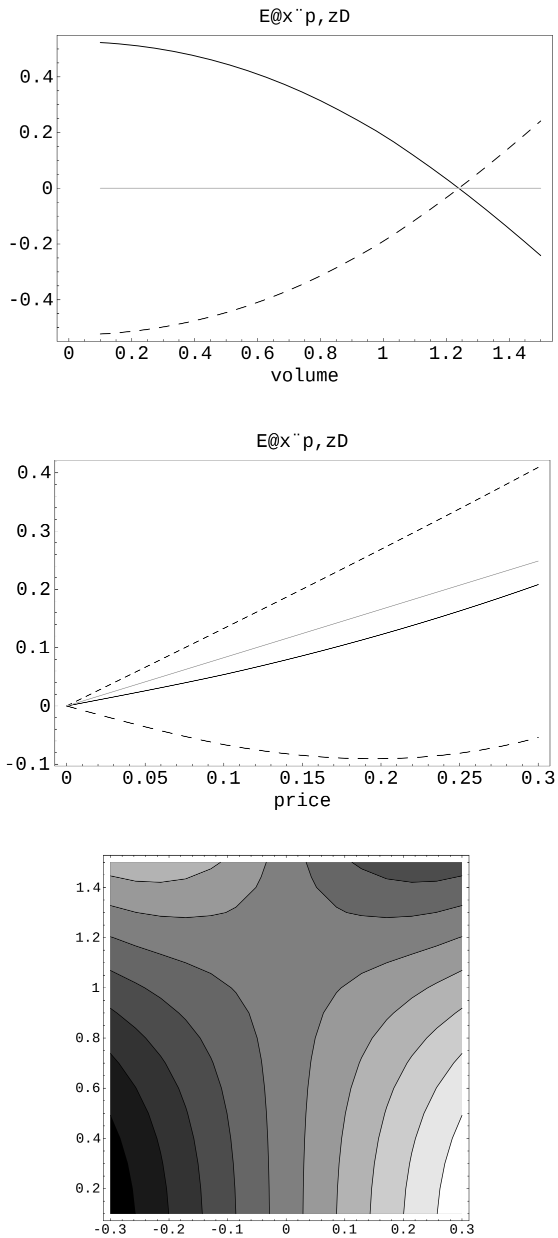

Figure 5: **Conditional mean conditioned on trading volume and price.** Upper panel: $E[x|p,z]$ as a function of z when $p = 0.5\sigma_p$ (solid line), $p = 0$ (gray line), and $p = -0.5\sigma_p$. Middle panel: $E[x|p,z]$ as a function of p when $z = E[z] - 0.5\sigma_z$ (short-dashed line), $z = E[z]$ (black solid line), $z = E[z] - 0.5\sigma_z$ (dashed line) and $E[x|p]$ (gray line) Lower panel: $E[x|p,z]$ as a function of p (horizontal axis) and z (vertical axis).

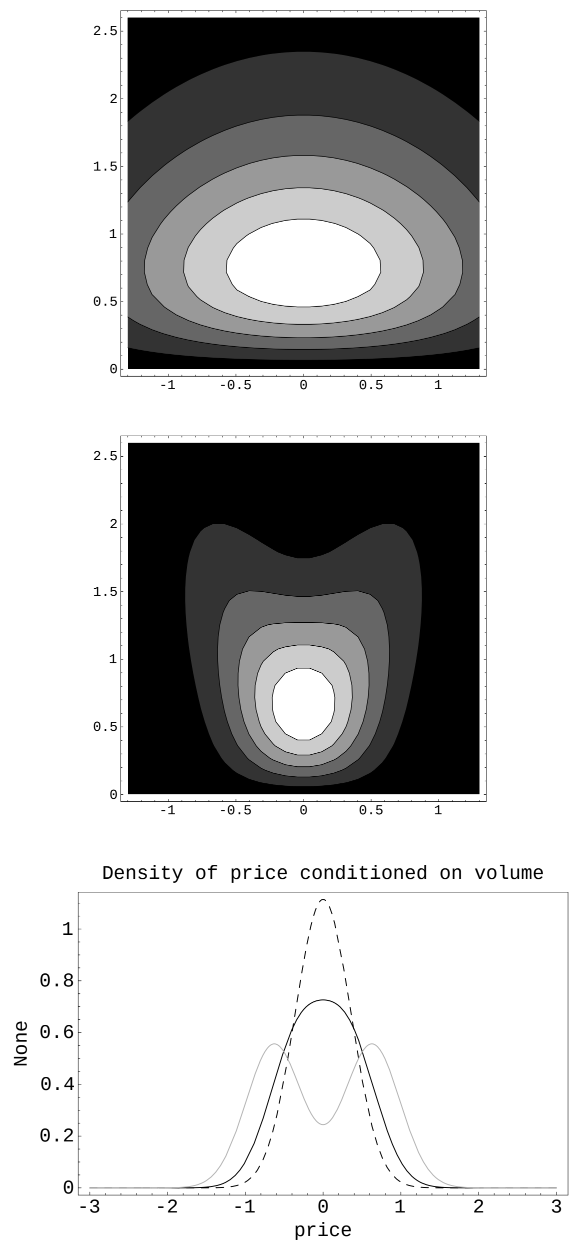

Figure 6: **Joint distribution trading volume and price or asset value.** Density dunctions of (x, z) (upper panel) and of (p, z) (middle panel). Conditonal density of x given z (lower panel) for $z = .25$ (dashed line), $z = 1.2$ (solid line), $z = 2$ (gray line).

REFERENCES

Blume Lawrence, David Easley, and Maureen O'Hara (1994), "Market Statistics and Technical Analysis: The Role of Volume", The Journal of Finance, Vol XLIX, No 1, March 1994

Diamond, Douglas and Robert E. Verrechia (1981), "Information Aggregation in a noisy rational expectation economy", Journal of Financial 9, p 221-235

Dupont, Dominique (1998), "Equilibrium price with institutional investors and with naive traders", working paper, Federal Reserve Board.

Foster, A. (1995), "Volume-Volatility Relationships for Crude Oil Futures Markets", The Journal of Futures Markets, vol 15, No 8, pp. 929-951.

Foster, A. and S. Viswanathan (1993), "The Implications of Public Information and Competition on Trading Volume and Price Volatility", Review of Financial Studies, vol 6, pp. 23-56.

Gallant, Ronald, Peter E. Rossi, and George Tauchen (1992), "Stock Prices and Volume", The Review of Financial Studies, Vol 5, No 2, pp 199-242

Goodman, William (1996), "Statistically Analyzing Volume", Technical Analysis of Stocks and Commodities, November 1996, p 21-28

He, Hua, and J. Wang, (1995) "Differential Information and the Behavior of Stock Trading Volume", The Review of Financial Studies, Vol 8, No 4, pp. 919-972.

Jain, P. and G. Joh (1988), "The Dependence between Hourly Prices and Trading Volume", Journal of Financial and Quantitative Analysis, Vol 23, No 3, September 1988, pp. 269-282.

Karpoff, Jonathan M. (1987), "The Relationship between Price Changes and Volume: A Survey", Journal of Financial and Quantitative Analysis, Vol 22, No 1, March 1987

Karpoff, J. M. (1988) "Costly Short Sales and the Correlation of Returns with Volume," The Journal of Financial Research, vol XI, No 3, pp. 173-188.

Kyle, Albert (1985), "Continuous Auction and Insider Trading", Econometrica, 53, p 1315-1335.

Wang, Jiang (1992), "A Model of Intertemporal Asset Prices under Asymmetric Information", The Review of Financial Studies, p 249-281

Wang, Jiang (1994), "A Model of Competitive Asset Trading Volume", Journal of Political Economy, vol 102, No 1, p 127-168