The Probability of Informed Trading: Applying EKOP to Canadian Data

Hausarbeit im Rahmen des Hauptseminars
Topics in International Economics & Finance (Master)
von Prof. Dr. Joachim Grammig
im Sommersemester 2007
Betreuerin: Franziska Peter

Löw Beer, David
Hauffstr. 23
72074 Tübingen
Internationale Volkswirtschaftslehre
6. Fachsemester
davidloewbeer@hotmail.com

Contents

The Probability of Informed Trading

DAVID LÖW BEER

Abstract

In this article I investigate the probability of informed trading. Using an empirical method developed by Easley et al. (1996), I estimate the risk of information-based trading for a small sample of Toronto Stock Exchange listed stocks. I use data on the cumulated transactions classified as buys and sells respectively to determine how frequently new information occurs, whether new information is good or bad news, and to estimate arrival rates for informed and uninformed traders. My estimates confirm the empirical results of Easley et al. (1996) in that information-based trading is more likely for less traded stocks. Furthermore, I hint at the fact that the probability of information-based trading seems to differ between industries.

1 Introduction

Apart from the intrinsic importance of having a sound knowledge about the probability of informed trading (PIT), the concept is also central to many fundamental questions in finance, among others: Does the PIT differ between more and less frequently traded stocks? Why do bid-ask spreads differ between securities? What kind of investor holds which stock? Why do companies split stocks?

Due to the great practical interest, various techniques of estimating the PIT have been developed in the history of financial markets econometrics. One possibility is presented by Hasbrouck (1991), who applies a vector autoregression model to quote and trade data. Thereby, he infers from trade innovations whether there is private information or not in the market and hopes to avoid misleading inferences due to inventory control or other transient liquidity effects. He concludes that, for stocks with small market values, trade innovations have greater persistent price impacts, which he interprets as arising from larger informational asymmetries, stating that market makers should take larger orders as a signal for an increased probability of informed trading.

The methodology developed by Easley et al. (1996) (hence EKOP) that will be presented in detail in the third section is different in various aspects. Firstly, it uses a likelihood function in order to estimate directly the market maker's beliefs. Secondly, information events may at the maximum occur only once per day.

Thirdly, it considers all order activities and not only those after trade innovations. Fourthly, the estimation relies only on cumulated sell and buy order arrivals and consequently needs fewer data.

The findings of the two estimating techniques do not differ widely, however the EKOP can explain the greater informational asymmetries of the less frequently traded stocks. Its empirical analysis shows that more frequently traded stocks display higher probability of information events and greater arrival rates of informed and uninformed traders. However, owing to the fact that the arrival rates of uninformed traders are so dramatically lower for less frequently traded stocks, the PIT is also lower for less active stocks. My empirical estimation of trade data from Canada confirms these estimates.

This paper is organized as follows. In the next Section, I present a literary review on some applications of EKOP. Then, there will be a summary of the structural model developed by Easley et al. (1996). Section 4 discusses the features of the Toronto Stock Exchange and the data. Results of the estimation are presented in Section 5. Finally, Section 6 concludes.

2   Literary Review

Since 1996, the EKOP has become a standard model to estimate the probability of informed trading. It has been applied and extended in many different analyses. Some of the results of these studies are presented in this section. Already in the original paper Easley et al. (1996) find that the higher probability of informed trading is a decisive factor in explaining the large bid-ask spreads that less frequently traded stocks normally display.

The work of Easley et al. (1998) investigates the role of financial analysts. The key finding is a negative correlation between the number of financial analysts trading with a stock and the probability of information-based trading. The paper shows that stocks with more analysts do involve more informed trade, but that they have even greater rates of uninformed trade. This suggests that while analysts' clients may be trading based on information, analysts attract even more uninformed traders to a stock. It is this greater depth that reduces the probability of information-based trading. Furthermore, their empirical evidence suggests that the probability of private information events is the same across stocks with many and few analysts.

Using EKOP, Easley et al. (2001) show that stock splits attract uninformed as well as informed traders. There is no overall significant effect in the information content of trades after a split. The paper, however, is able to show that trading costs rise after stock splits. This is due to an increase in volatility.

Easley et al. (2002) demonstrate that information-based trading has a large and significantly positive effect on asset returns. Using EKOP, it is possible to falsify traditional asset pricing models which assume that assets necessarily always include all information on the market. The results are robust, even when controlled for the correlation with spreads, variability in returns, and turnover. The work even reveals that the information variable and the firm size are the principal factors explaining returns.

Dennis et al. (2002) use EKOP to investigate the relation between ownership structure and informed trading. First, they find that the relative spread is negatively correlated to the amount of institutional ownership. This is attributed to the preference of institutions for stocks with narrower spreads since they are more liquid. Second, it is observed that information-based trading is significantly positively related to the amount of both institutional and inside ownership.

3 The EKOP Model

3.1 The Trading Process

Easley et al. (1996) have developed a mixed discrete and continuous time sequential trade model of market making. Within the framework individuals trade a single risky asset and money with a market maker over i = 1 , . . . , I trading days. Time within the trading day is indexed by $t \in [0,T]$.

Before a trading day starts, nature selects whether an information event relevant to the value of the observed asset occurs (with probability α) or not (with probability 1-α). Information events are independently distributed. These events are bad news with probability δ or good news with probability 1-δ.

Random variables $(V_i)_{i=1}^{I}$ give the value of the asset at the end of the day. If, on a given day, a bad event occurs, the value of the asset will be denoted $\underline{V_i}$, if a good event occurs $\overline{V_i}$, if no event occurs V_i^*.

Trade may involve informed (those who have observed any signal) and uninformed traders. Arrival rates of informed and uninformed traders are modelled by two in-

dependent Poisson processes [1] respectively. Uninformed buyers and sellers arrive at rate ε. They arrive on all trading days, independent of whether an information event has occurred or not. On days for which information events have occurred, informed traders arrive, too. The informed traders are assumed risk-neutral and competitive. In order to maximize her profits, the informed trader buys when observing a good signal and sells when observing a bad signal. All arrival processes are assumed to be independent and the arrival rate for each process is μ.
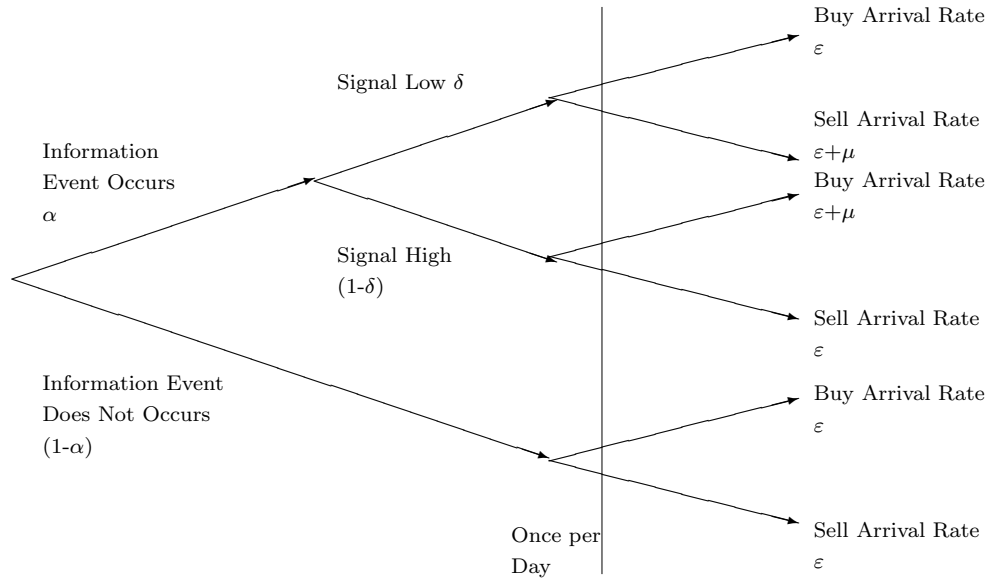

Figure 1. The structure of the trading process. The figure displays the trading process, where α is the probability of an information event, δ is the probability of "bad news", μ is the rate of informed trade arrival, and ε is the rate of uninformed trade arrivals. Nodes to the left of the vertical line occur once per day.

The tree given in Figure 1 illustrates this trading process. At the first node,

_____

[1] A Poisson distribution is defined as the number of the occurrences of a event including a great number (n→ ∞) of Bernoulli experiments, i.e experiments with only two possible results. A Poisson process is modeled by

$$P(x) = e^{-\lambda}\frac{\lambda^x}{X!}$$

nature determines whether an information event occurs. If an event occurs, it can be either good or bad news. There may be at the maximum one information event per day. Then, given the node selected for the day, traders arrive according to the Poisson process. That is, on a bad event day, trade arrival rates are ε for buy orders and $\varepsilon + \mu$ for sell orders. On a good event day, trade arrival rates are $\varepsilon + \mu$ for buy orders and ε for sell orders. On no event days only uninformed traders will arrive. Consequently, trades will arrive with rates ε for both buy and sell orders.

Prior to each trading day, the market maker knows the probabilities and the order arrival processes attached to each of the three branches. Since nature selects which of the three branches will be followed on a single day, the market maker does not know ex ante whether a specific day is a bad-event day, a good-event day or a non-event day. For information events are independent, the market makers prior beliefs about the probabilities of an information event and the trade arrival rates are equal for all days. Let $P(t) = (P_b(t), P_g(t), P_n(t))$ be the market makers prior belief about the events "bad news" (b), "good news"(g) and "no news" (n) at time t. So his prior belief at time 0 is $P(0) = (\alpha\delta, \alpha(1-\delta), 1-\alpha)$.

The EKOP model assumes a Bayesian [2] market maker who permanently revises his beliefs with respect to new information. So the bid and ask at any point in time t reflect both the history of the order process prior to the order arrival in t and the fact that someone wants to buy or sell. Let S_t indicate the event that a sell order arrives at time t and B_t the event that a buy order arrives. Let $P(t|B_t)$ be the market maker's updated belief vector conditional on the trade history prior to time t and on the event that someone wants to buy securities at time t. If an order to buy arrives at time t, the market makers posterior probability on bad

---

[2]The theorem of Bayes assumes that an individual revises his expectations because of new information. It establishes a connection between two conditional densities

P(A|B) and P(B|A)

Using the product rule for probabilities, this gives

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$$

Dividing both sides by P(B), providing that it is non-zero, it is straightforward to obtain Bayes' theorem in its most simple form:

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}$$

event is, by Bayes rule:

$$P_b(t|B_t) = \frac{P_b(t)\varepsilon}{\varepsilon + P_g(t)\mu} \tag{1}$$

Alike, the posterior probability on good news is

$$P_g(t|B_t) = \frac{P_g(t)\varepsilon + \mu}{\varepsilon + P_g(t)\mu} \tag{2}$$

Finally, the posterior probability on no news is

$$P_n(t|B_t) = \frac{P_n(t)\varepsilon}{\varepsilon + P_g(t)\mu} \tag{3}$$

It is straightforward to extend this to another moment in time when a sell order arrives.

Using these posterior probabilities, the competitive market maker sets his ask and bid prices at time t such that he expects profits to be zero. So, for the bid he will condition on the history prior to the order arrival and on the fact that someone wants to sell. Therefore the bid at time t on day i is

$$b(t) = \frac{P_b(t)(\varepsilon + \mu)\underline{V_i} + P_g(t)\varepsilon\overline{V_i} + P_n(t)\varepsilon V_i^*}{\varepsilon + P_b(t)\mu} \tag{4}$$

Similarly, the ask is

$$a(t) = \frac{P_b(t)\varepsilon\underline{V_i} + P_g(t)(\varepsilon + \mu)\overline{V_i} + P_n(t)\varepsilon V_i^*}{\varepsilon + P_g(t)\mu} \tag{5}$$

The expected value of an asset conditional on the trade history is, again by Bayes rule:

$$E[V_i|t] = P_b(t)\underline{V_i} + P_g(t)\overline{V_i} + P_n(t)V_i^* \tag{6}$$

Substituting equation (6) into the bid and ask equations (4) and (5), respectively, yields

$$b(t) = E[V_i|t] - \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)}(E[V_i|t] - \underline{V_i}) \tag{7}$$

and

$$a(t) = E[V_i|t] - \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)}(\overline{V_i} - E[V_i|t]) \tag{8}$$

Equations (7) and (8) show the decisive role of informed and uninformed trade arrivals when determining trading prices. If there are no informed traders (μ=0), then trading prices equal the prior expected value of the asset because trading does not convey any information. If there are only informed traders (ε=0), then one obtains b(t) = $\underline{V_i}$ and a(t) = $\overline{V_i}$. Normally, both informed and uninformed traders will be in the market. So the bid is below $E[V_i|t]$ and the ask is above $E[V_i|t]$. The spread that the market maker sets rises with a growing share of informed traders, because a larger share of informed traders signify greater risks of losses to informed traders. The resulting spread s(t) can be written explicitly by

$$s(t) = \frac{\mu(P_g(t))}{\varepsilon + \mu P_g(t)}(\overline{V_i} - E[V_i|t]) + \frac{\mu(P_b(t))}{\varepsilon + \mu P_b(t)}(E[V_i|t] - \underline{V_i}) \tag{9}$$

The spread at time t is the probability that the buy is information-based times the expected loss to an informed buyer, plus a symmetric term for sells. Therefore the probability that any trade occurring at time t is information based is the sum of these probabilities

$$PIT(t) = \frac{\mu(1 - P_n(t))}{\mu(1 - P_n(t)) + 2\varepsilon} \tag{10}$$

At the opening it is reasonable to assume that a good news event and a bad news event are equally likely. So, we have

$$PIT \equiv PIT(0)$$

$$= \frac{\mu(\alpha\delta + \alpha - \alpha\delta)}{2\varepsilon + \mu(\alpha\delta + \alpha - \alpha\delta)}[\overline{V_i} - \underline{V_i}]$$

$$= \frac{\alpha\mu}{\alpha\mu + 2\varepsilon}[\overline{V_i} - \underline{V_i}] \tag{11}$$

using the unconditional probabilities.

3.2 The Likelihood Function

Easley et al. (1996) constructed a structural model to estimate the relevant parameter vector $\theta = (\alpha, \delta, \varepsilon, \mu)$. The difficulty of this estimation lies in the fact that the outsider can only observe buys and sells. It is impossible to know which traders are informed and which are uninformed. Furthermore, we do not know

whether an information event has occurred and if so, whether it is bad or good news. However, since the data reflect the underlying information structure, it is possible to estimate ex post whether an information event has occurred or not. For example, if we assume a good news event day, then buy orders arrive at a rate $(\varepsilon + \mu)$ because both informed and uninformed traders arrive, whereas sell orders are at a rate ε only, reflecting that only uninformed traders will sell on a good-event day. Thus, the likelihood of observing any sequence of orders that contains B buys and S sells on a good-event day of total time T is given by

$$e^{-(\varepsilon+\mu)T}\frac{[(\varepsilon + \mu)T)]^B}{B!}e^{-\varepsilon T}\frac{(\varepsilon T)^S}{S!} \tag{12}$$

Analogically, on a bad-event day both uninformed and informed trader will sell the stock, but only uninformed traders will buy the stock. Accordingly, the likelihood of observing any order sequence on a bad-event day is given by

$$e^{-\varepsilon T}\frac{(\varepsilon T)^B}{B!}e^{-(\varepsilon+\mu T)}\frac{[(\varepsilon + \mu)T]^S}{S!} \tag{13}$$

Finally, on a no-event day, only uninformed traders will buy and sell stocks, since there is no incentive for an informed trader to get to the market. Therefore, the likelihood of observing an order sequence with B buys and S sells on a no-event day is

$$e^{-\varepsilon T}\frac{(\varepsilon T)^B}{B!}e^{-\varepsilon T}\frac{(\varepsilon T)^S}{S!} \tag{14}$$

Equations (12), (13), and (14) show that it is only necessary to consider the number of buys and sells in order to estimate the order arrival rates. Using these three equations and weighting them with the probabilities of good-event days $\alpha(1-\delta)$, bad-event days $\alpha\delta$, and no-event days $1-\alpha$ respectively, yields the likelihood

$$
\begin{aligned}
L((B,S)|\theta) = {} & (1 - \alpha) \cdot e^{-\varepsilon T}\frac{(\varepsilon T)^B}{B!}e^{-\varepsilon T}\frac{(\varepsilon T)^S}{S!} \\
& + \alpha\delta \cdot e^{-\varepsilon T}\frac{(\varepsilon T)^B}{B!}e^{-(\varepsilon+\mu)T}\frac{[(\varepsilon + \mu)T]^S}{S!} \\
& + \alpha(1-\delta) \cdot e^{-(\varepsilon+\mu)T}\frac{[(\varepsilon + \mu)T]^B}{B!}e^{-\varepsilon T}\frac{(\varepsilon T)^S}{S!}
\end{aligned} \tag{15}
$$

On the assumption that only one information event can occur per day the information event parameters α and δ will either be 0 or 1. Whether an event occurred or not can be estimated from the daily number of buys and sells. Because information events are independent, the likelihood of observing the data M=(B_i,S_i),

which is the joint distribution of the daily data, over I days is just the product of the individual densities given in the previous equation,

$$L(M|\theta) = \prod_{i=1}^{I} L(\theta|B_i, S_i) \tag{16}$$

To estimate the parameter vector θ from any data set M, it is necessary to maximize this likelihood. This provides direct estimates of the rate of informed and uninformed trading in a particular asset, as well as of the information event structure surrounding that asset.

4   The Data

Fifteen stocks were randomly selected from the 1,500 stocks traded on the Toronto Stock Exchange (hereafter TSX), the world's 7th largest Stock Exchange with a market capitalization of 1.75 trillion US dollars. TSX is the global leader for listing mining, and oil and gas companies operating around the world.

Stocks in Toronto are traded electronically only. The role of the Market Maker on TSX is that of a passive liquidity manager. The idea is that the normal trading process is organized via a continuous auction market. Orders are organized by a computer-based central limit order book to which all market participants have full access. Hence, the market seems to be highly transparent, however, anonymous trades are possible and there are mechanisms to disguise "large" orders. Following Grammig et. al (2001) who compare the probability of informed trading between the non-anonymous traditional floor based exchange and the anonymous computerized trading system and due to the high amount of international capital on the TSX, I expect relatively high rates of informed trading.

For estimating the parameters trade data for the 15 stocks in my sample I consider the first trading hour of the 64 trading days between 01/01/2004 and 31/03/2004. There was one technical and one practical reason for selecting the first hour of trading. From the technical perspective, it was necessary to restrict the number of buys and sells to limit the factorial needed for the calculation of the likelihood given in equation (16). Therefore, it seemed quite reasonable to select the first trading hour, as it is the time when prices adjust after the non-trading period during the night.

Following Easley et al. (1993) the period of 64 days is considered at the same

time long enough to estimate the parameters correctly and short enough so that the stationarity built in the trade model is not too unreasonable.

The list of the selected stocks, their respective industries, their average cumulated number of buy and sell transactions in the first trading hour, and their average prices during the first trading hour is provided in the Appendix (see Table A.1.)

To compute the likelihood function given in equation (16) and to consequently determine the PIT, it is necessary to estimate the number of sells and buys on each day for each stock.

Trades were classified using the algorithm developed by Lee and Ready (1991) If the price is above (below) the prevailing midpoint quote, the trade is classified as a buy (sell). Trades at midquote are called buys (sells) if the price is higher (lower) than the price of the most recent trade. This procedure is standard. However as Boehmer et al. (2007) show, the methodology misclassifies so many trades that it causes a downward-bias of the PIT. The bias is particularly large for less frequently traded assets.

Finally, I ranked all stocks by the sum of cumulated sells and buys. Since I did not have access to the trading volume data, these numbers are taken as a proxy. However, this approach is problematic, because each trade may include the selling and buying of various stocks.

5   Estimation

In this section, I will present the results of my parameter estimates for the structural model. Recall that the trade process depends on the parameter vector θ, which combines four parameters: α is the probability of an information event; δ the probability that an information is bad news; ε the arrival rate of uninformed traders; and μ the arrival rate of informed traders. Using the estimates, it is possible to calculate the probability of informed trading applying equation (10).

5.1 Parameter Estimates

The trade data for each stock in our sample is estimated by maximizing the likelihood which was developed in the previous section. All four parameters were restricted to be non-negative, the information parameters α and δ were additionally restricted to (0,1).

I maximize the unrestricted parameter estimates using various algorithms in the GAUSS statistical package. Thereby the computer tests possible values for θ and chooses the value that make the likelihood of the observed data largest. Put differently, it chooses the value for θ, which, if it were the true parameter of the distribution, would generate the sample results with the greatest probability compared to all other possible values.

Parameter estimates and their robust standard errors for each stock are provided in the Appendix (see Table A.2). The standard errors show that the model can be estimated very precisely. Trade arrival rates ε and μ for all assets are significant on any convenient significance level. The information parameters α and δ are significant on the 1% significance level for all but one stock in the sample.

Table 1 provides the means, the medians and the mean standard errors of the estimated parameters. Since EKOP was developed to evaluate differences between less and more active traded stocks, I present the results divided by three terciles, the first includes the stocks with the greatest number of cumulated buys and sells in the first trading hour, the second those with the 6th to 10th greatest number of cumulated buys and sells in the first sixty minutes of trading, and the third tercile is composed of the five assets of my sample, which display the smallest number of cumulated buys and sells in the first trading hour of each day.

The parameter α is the probability of an information event to occur before the opening of the stock exchange. It can take either the value 0 or 1. Table 1 shows that the mean α is 0.327 for all stocks, 0.460 for the first tercile, 0.275 for the second tercile, and 0.246 for the third tercile. Thus, the probability of information events is highest for the most active stocks in the sample, and declines for the less active ones. Cumulative distributions of α tend to differ across terciles, with that of the most traded stocks higher than those of the less frequently traded stocks, and still higher than the least traded stocks. [3] The original EKOP paper finds similar results.

---

[3] However the data for all three categories show great variability. For the most active stocks, α ranges between 0.29 and 0.63 and for the least active ones between 0.06 and 0.57. Thus, there are infrequently traded stocks that often have new information, and there are frequently traded stocks that rarely have new information.

Table 1. Summary Parameter Estimate Statistics

This table presents means, medians, and mean sample standard deviations of parameter estimates by volume tercile for the 15 stocks in my sample. The parameter α is the probability of an information event, δ is the probability that new information is bad news, ε is the arrival rate of uninformed traders, and μ is the arrival rate of informed traders. The parameter PIT is a composite variable measuring the probability of information-based trade.

Parameter	All	First Tertile	Second Tertile	Third Tertile
Number in Sample	15	5	5	5
α				
Mean	0.327077	0.459981	0.274766	0.246484
Median	0.284337	0.461272	0.272253	0.119486
Std. dev.	0.005127	0.004421	0.004272	0.006687
δ				
Mean	0.237371	0.102034	0.311893	0.298186
Median	0.233881	0.085091	0.257103	0.359449
Std. dev.	0.281361	0.003799	0.010539	0.829744
ε				
Mean	0.809126	1.583830	0.622800	0.220749
Median	0.711807	1.454631	0.459114	0.221676
Std. dev.	0.000084	0.000203	0.000030	0.000019
μ				
Mean	1.178585	1.417456	1.152873	0.965427
Median	1.000072	1.122029	1.207071	0.671116
Std. dev.	0.001143	0.001652	0.001412	0.000365
PIT				
Mean	0.195122	0.163455	0.200759	0.221153
Median	0.182157	0.151032	0.184949	0.210491
Std. dev.	0.009096	0.000573	0.001906	0.024810

The other information parameter in our model is δ, which measures the probability that a new information is bad news. It can also take either the value 0 or 1. From a theoretical perspective there is no reason to suspect any differences

in the probability of bad news with regards to the size of a stock. Empirically, Easley et al. (1996) cannot reject the hypothesis of no differences in the three distributions. In my estimation, the δs are 0.102 for the five most traded stocks in our sample, 0.312 for the five second most traded stocks, and 0.298 for the five least traded stocks. The average δ is 0.237. So, although I did not test whether those differences are significant, a superficial first look, indicates this. Yet, the differences might be due to the small sample size. It should also be noted that the cumulative distributions exhibit multiple crossings. So, there is only very limited evidence against the theoretically posted hypothesis of δs, which do not differ across the terciles.

I turn now to the arrival rates of informed and uninformed traders. Since these are intensities and not probabilities, they could only be interpreted if they were multiplied by a small period of time, ΔT. Both rates are only restricted to be non-negative.

Table 1 shows huge differences in the arrival rates of uninformed traders, ε. The estimated mean for all stocks is 0.809. It is 1.583 for tercile 1, 0.622 for tercile 2, and 0.221 for tercile 3. Comparing the cumulative distributions within each tercile yields the same results, i.e. all of the more active stocks display much higher εs than the less active stocks. This is consistent with the analysis of Easley et. al (1996), also showing that more active stocks attract more uninformed traders.

The results for informed order arrivals, μ, are quite similar, although the differences are less dramatic. The overall rate is 1.179. It falls from 1.417 for the first tercile to 1.153 for the second tercile to 0.965 for the third tercile. Although showing some crossings, distributions of δ tend to differ across the terciles being higher for the more actively traded stocks and lower for the less actively traded securities. Conforming again the results of the original EKOP paper, I find that the informed arrival rate is higher for more active stocks. [4]

5.2 The Probability of Informed Trading

As it is shown in equation (10) the probability of informed trade is a composite variable reflecting the probability that new information exists and the trade arrival

---

[4]Notice, however, that, contrary to the data for uninformed traders, ε, the data for the arrival rate of informed traders, μ, display great variance. The stock KFS, e.g., which is classified as one of the five least traded stocks in the sample shows the highest trade arrival rate of informed buyers and sellers.

parameters. Table 1 shows that the estimated mean of PIT is 0.195 for all traded stocks, 0.163 for the five most active stocks, 0.201 for the five second most active stocks and 0.221 for the five least active stocks. Thus the risk of informed trading is negatively correlated with the activity of a stock. For the more frequently traded stocks, the higher probability of information events and the higher arrival rates of informed traders are more than offset by the dramatically higher arrival rates of uninformed traders. The results are consistent with the findings of Easley et al. (1996).

5.3 Probability of Informed Trading and Industry Sector

The Toronto Stock Exchange classifies five of the fifteen randomly selected companies in the industrial category "mining", five in "industrial", three in "oil/gas", and one in "life sciences" and "communications" respectively. Table 2 provides summary statistics of the probability of informed trading grouped by the respective industries.

Table 2. The Probability of Informed Trading and Industry Sector.
Standard errors are given in parenthesis.

Stocks in Category	Industry	PIT
5	Mining	0.18772
		(0.00083)
5	Industrial	0.18278
		(0.02453)
3	Oil/Gas	0.23220
		(0.00279)
2	Others	0.18886
		(0.00065)

Although the data base is far too small to make a serious statement, and even though my knowledge on the ownership structure of the companies in the sample is far too little to explain possible differences, a first glimpse at the empirical results shows that this approach might be interesting for further research. Companies trading with natural ressources might, in contrast to industrial firms, attract less uninformed traders because of the presumably higher volatility or due to the fact

that small, local investors are more likely to buy stocks of less internationalized companies.

6   Conclusion

Applying the empirical technique developed by Easley et al. (1996), I use data from the Toronto Stock Exchange to estimate the probability of informed trading for fifteen stocks. I divide the stocks into three groups with respect to the cumulated number of transactions classified as buys and sells. This is used as a proxy for trading volumes.

Conforming the results of Easley et al. (1996), I also find that the probability of information-based trading is lower for more frequently traded stocks. Even though less frequently traded stocks have lower probabilities of information events and lower arrival rates of informed traders, these are more than counterbalanced by the much lower arrival rates of uninformed traders.

Using the estimates provided by EKOP, it is possible to answer some of the core questions in financial market econometrics. The data reveal that the bid-ask spread is larger for less active stocks, because the ratio of informed to uninformed traders is higher in those assets. Furthermore, it is possible to show that the more financial analysts trade with a stock the lower is the probability of informed trading.

The probability of informed trading (PIT) of the stocks in my sample differs greatly between industries. Particularly the assets from the oil/gas sector show much higher PIT than those from the mining and industrial sector. This could be due to a great variety of reasons, among others the structure of investors might differ across industries. However, these questions have to be left for further research.

7 Appendix

Table A.1
TSX Stocks Included in Sample

This table presents the stocks included in the testing of EKOP. Fifteen stocks of the Toronto Stock Exchange were randomly selected. The firms were classified into different industries with reference to the company information provided by the TSX homepage. Data on the cumulated transactions classified as buys (Buys) and sells (Sells) respectively and on the average price (AvgPrice) of the stock refer to the first trading hour of the 64 trading days between 01/01/2004 and 03/31/2004. Average Price is mean price of the first trading hour. Data is ranked by the sum of cumulated buys and sells in the first trading hour.

Ticker	Company Name	Industry	Buys	Sells	Price
AL	Alcan Inc.	Industrial	157.78	125.34	59.67
BCE	Bell Canada Enterprises Inc.	Communications	158.14	121.06	28.93
ABX	Barrick Gold Corp.	Mining	115.55	90.06	27.94
MFC	Manulife Financial Corporation	Industrial	109.53	75.72	46.63
PCA	Petro-Canada	Oil/Gas	112.2	62.31	60.88
NRD	Noranda Inc	Mining	71.42	88.91	21.17
NXY	Nexen Inc.	Oil/Gas	53.69	44.11	49.27
PKZ	PetroKazakhstan Inc.	Oil/Gas	43.97	25.7	34.06
AGE	Agnico-Eagle Mines Ltd.	Mining	35.61	33.05	17.76
GLG	Glamis Gold Ltd.	Mining	36.91	29.61	21.41
KFS	Kingsway Financial Services Inc.	Industrial	31.03	18.41	14.62
AGU	Agrium Inc.	Mining	26.06	20.7	19.81
MDS	MDS Inc.	Life Sciences	18.25	16.64	20.99
MWI	Moore Wallace Incorporated	Industrial	11.69	11.17	25.01
IPS	IPSCO Inc.	Industrial	10.34	6.58	23.09

Table A.2

Continous Time Trading Model Parameter Estimates

This table presents the parameters estimated using EKOP. The parameter μ is the arrival rate of informed traders, ε is the arrival rate of uninformed traders, α is the probability of an information event, and δ is the probability that new information is bad news. The parameter PIT is a composite variable measuring the probability of information based-trade. Standard errors are given in parenthesis below the parameter estimates. Maximum likelihood estimation is performed using various algorithms in the GAUSS statistical package.

Ticker	μ	ε	α	δ	PIT
AL	0.337893	2.064990	0.076956	1.742477	0.124772
	(0.003922)	(0.000430)	(0.004222)	(0.005347)	(0.000400)
BCE	0.631149	2.000627	0.085091	1.033258	0.140143
	(0.004421)	(0.000484)	(0.002211)	(0.001842)	(0.000181)
ABX	0.461272	1.454631	0.153472	1.122029	0.151032
	(0.004839)	(0.000040)	(0.004922)	(0.000629)	(0.000409)
MFC	0,576212	1,276310	0,132526	0,928264	0,173240
	(0.004684)	(0.000046)	(0.003937)	(0.000323)	(0.000319)
PCA	0.293382	1.122592	0.062122	2.261251	0.228086
	(0.004197)	(0.000015)	(0.003701)	(0.000119)	(0.001555)
NRD	0.352981	1.116504	0.686416	1.244038	0.164334
	(0.005332)	(0.000027)	(0.011825)	(0.000211)	(0.000850)
NXY	0.249743	0.664245	0.257103	1.207071	0.184949
	(0.005755)	(0.000017)	(0.012630)	(0.000327)	(0.002290)
PKZ	0.171866	0.415959	0.090914	1.915889	0.283567
	(0.003147)	(0.000005)	(0.007742)	(0.000083)	(0.004520)
AGE	0.326988	0.459114	0.346905	0.691280	0.197541
	(0.003804)	(0.000022)	(0.011812)	(0.001316)	(0.000921)
GLG	0.272253	0.458181	0.178127	0.706087	0.173403
	(0.003323)	(0.000077)	(0.008686)	(0.001524)	(0.000948)
KFS	0.062501	0.327554	0.000102	2.701608	0.204928
	(0.015453)	(0.000004)	(4.006382)	(0.000393)	(0.105295)
AGU	0.401124	0.291381	0.296520	0.490261	0.252311
	(0.004264)	(0.000057)	(0.008585)	(0.000709)	(0.001019)
MDS	0.571030	0.221676	0.435882	0.241946	0.237586
	(0.007885)	(0.000022)	(0.008241)	(0.000181)	(0.001122)
MWI	0.119486	0.150386	0.398978	0.671116	0.210491
	(0.003976)	(0.000007)	(0.060035)	(0.000326)	(0.008317)
IPS	0.078278	0.112748	0.359449	0.722206	0.200452
	(0.001857)	(0.000004)	(0.065476)	(0.000217)	(0.008296)

References

[1] Boehmer, E., Grammig, J., and E. Theissen, 2007. Estimating the probability of informed trading — does misclassification matter? Journal of Financial Markets 10, 26-47

[2] Dennis, P., and J. Weston, 2001: Who's Informed? An Analysis of Stock Ownership and Informed Trading. Journal of Financial Research, 439-459

[3] Easley, D., Kiefer, N., and M. O'Hara, 1997. One day in the life of a very common stock. Review of Financial Studies 10, 805-835

[4] Easley, D., Kiefer, N., O'Hara, M., and J. Paperman, 1996. Liquidity, information, and infrequently traded stocks. Journal of Finance 51, 1405-1436

[5] Easley, D., and M. O'Hara, 1987. Price, trade size and information in security markets. Journal of Financial Economics 19, 69-90

[6] Easley, D., and M. O'Hara, 1992. Time and the process of security price adjustment. Journal of Finance 47, 577-606

[7] Easley, D., O'Hara, and J. Paperman, 1998. Financial analysts and information-based trade. Journal of Financial Markets 1, 175-201.

[8] Easley, D., Hvidkjaer, S., and M. O'Hara, 2002. Is information risk a determinant of asset returns? Journal of Finance 57, 2185-2221.

[9] Easley, D., O'Hara, M., and G. Saar, 2001. How stock splits affect trading: A microstructure approach. Journal of Financial and Quantitative Analysis 36, 25-51.

[10] Grammig, J., Schiereck, D., and E. Theissen, 2001. Knowing me, knowing you: Trader anonymity and informed trading in parallel markets. Journal of Financial Markets 4, 385-412

[11] Hasbrouck, J., 1991. Measuring the Information Content of Stock Trades. The Journal of Finance 46, 179-207