**Informed trading, market efficiency and volatility**

Ming-Chien Sung\*1, Johnnie E. V. Johnson1, David C. J. McDonald1

1 Centre for Risk Research, University of Southampton, Southampton, SO17 1BJ, UK.

(Tel: +44(0) 23 80598974; +44(0) 23 80592546; +44(0) 23 80599248. Email: [m.sung@soton.ac.uk](mailto:m.sung@soton.ac.uk); [jej@soton.ac.uk](mailto:jej@soton.ac.uk); [d.mcdonald@soton.ac.uk](mailto:d.mcdonald@soton.ac.uk))

\*corresponding author

**Abstract**

We establish relationships that have proved difficult to capture in financial markets, between informed trading, efficiency and volatility. We examine the efficiency and volatility of market prices in 6058 parallel horserace betting exchange and bookmaker markets (1.8 million price points). We find that informed trading is associated with increased efficiency and volatility.

**Key words**: informed trading, efficiency, volatility

**JEL classification**: G12, G14

**1. Introduction**

We examine relationships between informed trading and two important market features: efficiency, the predictive accuracy of prices with respect to underlying fundamentals, and volatility, variations in market prices around fundamentals. Financial market studies provide only limited evidence concerning these relationships.

New information has an unpredictable effect on share prices (Johnstone, 2016). However, it might be expected that informed trading is positively associated with market efficiency (e.g., Tavakoli et al., 2012). Uninformed trading introduces risks for risk-averse informed traders, encouraging inefficiency (Shiller, 1990; e.g., markets may remain inefficient longer than informed traders can remain liquid, causing them to limit arbitrage). An alternative possibility is observed in Bloomfield et al.’s (2009) experimental market: provided prices are not extreme, noise makes prices *more* efficient by providing liquidity.

The relationship between informed trading and volatility is less clear. French and Roll (1986) find that greater volatility in trading (cf. non-trading) hours can only be explained by private information (cf. noise). However, De Long et al. (1990) predict that, in the absence of new information, uninformed trading increases short-term volatility. Previous tests have proved inconclusive.

We shed light on these relationships using a dataset valuable for exploring the role of informed trading: horserace betting market data. These markets share many characteristics with wider financial markets, incorporate a defined end-point at which all uncertainty is resolved, and there are many thousand independent markets each year (Sauer, 1998; Vaughan Williams and Paton, 1997; Law and Peel, 2002).

To measure informed trading, we employ Shin *z* (Shin, 1993), which measures the proportion of market participation by traders with privileged information (not necessarily insider trading, e.g., Peirson and Smith, 2010). Shin’s game-based model of a horseracing market consists of an expected profit-maximizing bookmaker and a randomly selected bettor who is either perfectly informed (i.e., know the winning horse in advance) or a noise trader. The model predicts that the bookmaker will depress odds relatively more on longshots as protection from the informed trader. Shin *z* is derived from closing bookmaker prices and has been used extensively to investigate claims relating to informed trading (e.g., Vaughan Williams and Paton 1997; Smith et al., 2006).

The hypotheses we test are that *informed trading is associated with H1: increased market efficiency*, and H2: *increased* *market price volatility*.

**2. Data and method**

*2.1 Data*

The data are drawn from parallel bookmaker and exchange markets for 62,124 horses running in 6058 races in the UK and Ireland between August 2009 and August 2010. The bookmaker prices (used to estimate Shin *z*) are the mean closing prices offered by a broad cross-section of bookmakers, while the exchange prices are a time-series of the best available prices from the largest exchange market (Betfair) at 1-minute intervals throughout the last 30 minutes of each market (when volumes are greatest): 1.8 million data points.

We consider ‘win’ markets, in which bettors must predict which horse will win (or lose). The odds for horse *i* in race *j* is the return to a $1 winning bet (e.g., a winning $1 bet with odds of 3.00 returns $3 (a profit of $2)). Betfair take commission of 5% from winnings, so effective odds are given by . The odds-implied probability *qij* of horse *i* winning race *j*, with *nj* runners, is given by

(1) .

Each race *j* results in a vector of outcomes , where *yhj* = 1 for the winning horse *h* and *yij* = 0 otherwise. If markets are efficient, then, over many races, odds-implied probabilities should approximate realized race results.

*2.2 The conditional logit model and efficiency*

We estimate Shin *z* using Law and Peel’s (2002) iterative method based on closing bookmaker odds. We assume that, given their parallel nature, higher levels of informed activity in the bookmaker market are reflected in the exchange market. We compare the efficiency of final exchange prices in races with above and below-median Shin *z* (3029 races in each of the ‘informed’ and ‘uninformed’ sets). Specifically, we estimate single-variable conditional logit (CL) models (McFadden, 1974) on each set of races, where the probability of competitor *i* winning race *j* is given by

(2) ,

The coefficient *β* is estimated by maximum likelihood:

(3) .

Here, *yij* = 1 if horse *i* won race *j* and *yij* = 0 otherwise, and *N* is the total number of races in the dataset. An appropriate measure of the predictive accuracy of the model, and therefore market efficiency, is Maddala’s (1983) pseudo-*R*2, given by

(4) ,

where ln*L*0 is the log-likelihood of the naive model (each horse in a race is assigned the same probability of winning):

(5) .

An estimated value of *β* = 1 implies that there is no favorite-longshot bias (FLB: the widely-reported phenomenon whereby favorites/longshots are under-/over-bet, e.g., Gramm and Owens, 2006).

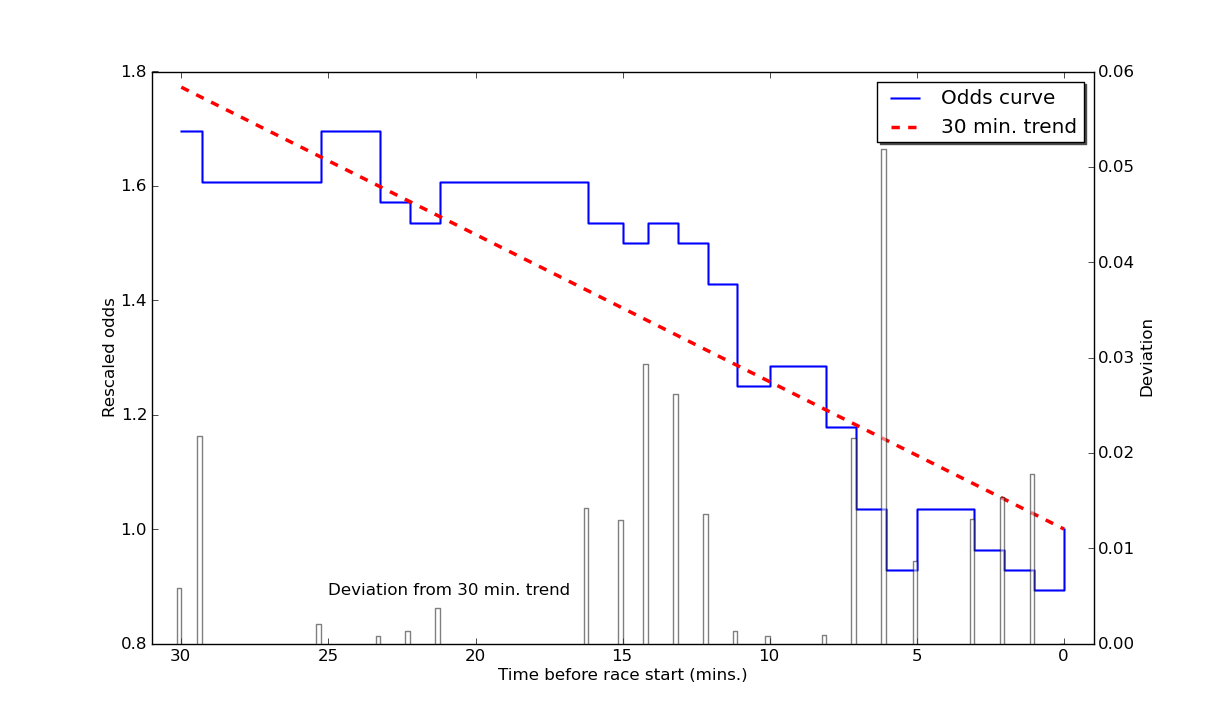
We estimate the distributional properties of the pseudo-*R*2s using a bootstrap method. For each of the informed/uninformed subsets, we repeat 1000 times a random sampling of 3029 races, with replacement, and fit a model. The sample means, and , and variances, and , of the pseudo-*R*2s from the informed/uninformed subsets, respectively, result in a standard normal test statistic to test H1:

(6) .

*2.3 Volatility*

Information that is revealed during the course of the market helps determine accurate winning probabilities. Consequently, we expect, *ex ante*, that prices will fluctuate around an underlying trend, representing the bettors’ collective opinion of the horse’s chances at the market close relative to the opening. We calculate, for each horse, in each race, the trend *μij* in the odds curve in the exchange market using the method of Johnson et al. (2006). The volatility *σij* is given by the variance of the odds around the trend. To illustrate, Figure 1 shows an example of an odds curve with the 30 minute trend, along with the deviation. For the overall volatility in the market, we use the mean of *σij* over all the horses in each race, . This allows us to test H2.

Figure 1. Example of the least squares regression method for determining the trend and volatility of a horse’s odds curve.



**3. Results**

The median Shin *z* over all the data is 0.0198, and the informed/uninformed trading sets have mean Shin *z* of 0.0259 and 0.0172, respectively. Further characteristics of the data in each set and their 1000-bootstrapped samples, are summarized in Table 1.

Table 1. Details of efficiency and volatility of standard and bootstrapped datasets with above- and below-median levels of informed trading.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | All data | ‘Informed’ trading half | ‘Uninformed’ trading  half |  | Bootstrapped ‘informed’ trading half | Bootstrapped ‘uninformed’ trading half |
| No. of races | 6058 | 3029 | 3029 | No. of races | 3029 | 3029 |
| No. of horses | 62,124 | 29,038 | 33,086 | Mean no. of horses | 29,153 | 32,924 |
| Mean no. of horses per race | 10.3 | 9.6 | 10.9 | Mean no. of horses per race | 9.6 | 10.9 |
| No. of handicaps | 3255 | 1361 | 1894 | Mean no. of handicaps | 1628.0 | 1626.8 |
| Proportion of handicaps | 0.537 | 0.449 | 0.625 | Mean proportion of handicaps | 0.537 | 0.537 |
| Mean Shin *z* | 0.0215 | 0.0259 | 0.0172 | Mean Shin *z* | 0.0256 | 0.0172 |
| Level of FLB *β* | 1.014 | 1.017 | 1.010 | Mean level of FLB *β* | 1.022 | 1.013 |
|  |  |  |  | *z*(*β*)  (*S.E.*) | 0.26  (0.034) | |
| [*β*-1] / *S.E.*[*β*]    (*S.E.*[*β*]) | 0.78  (0.017) | 0.71  (0.024) | 0.39  (0.025) | [*β*-1] / *S.E.*[*β*]  (*S.E.*[*β*]) | 0.93  (0.023) | 0.52  (0.025) |
| ln*L*(naive) | -13,670 | -6555 | -7116 | Mean ln*L*(naive) | -6571 | -7100 |
| ln*L*(full) | -11,125 | -5132 | -5993 | Mean ln*L*(full) | -5229 | -5894 |
| Efficiency  (pseudo-*R*2) | 0.5684 | 0.6091 | 0.5235 | *μ*(*R*2) | 0.5875 | 0.5489 |
|  |  |  |  | *z*[*μ*(*R*2)]  (*S.E.*) | 4.77\*\*  (0.018) |  |
| Mean volatility | 0.0242 | 0.0270 | 0.0214 | Mean volatility | 0.0260 | 0.0222 |
|  |  |  |  | (*S.E.*) | 5.79\*\*  (0.001) |  |

\* and \*\* denote significance at the 5% and 1% level in a 2-tailed test, respectively.

CL models, with ln*qij* as the single predictor variable, are estimated for the two sets, as well as for each of the bootstrap samples. The coefficient *β* is not significantly different from 1 in any case (*z* = 0.71, *p* = 0.4770, and *z* = 0.39, *p* = 0.6982 for the original informed/uninformed sets, respectively; *z* = 0.93, *p* = 0.3502, and *z* = 0.52, *p* = 0.6049 for the bootstrapped sets), indicating, as expected, the absence of FLB (Smith et al., 2006). Furthermore, we find that when Shin *z* is higher, prices are on average more accurate in predicting winners (mean pseudo-*R*2 = 0.5875 and 0.5489 for informed and uninformed, respectively; *z* = 4.77, *p* = 0.0000) and more volatile (mean volatility 0.0260 and 0.0222 for informed and uninformed, respectively; *z* = 5.79, *p* = 0.0000) (see Figures 2 and 3). Consequently, the results support our hypotheses that increased informed trading is associated with increased market efficiency and volatility.

Figure 2. Histograms of pseudo-*R*2 from 1000-bootstrap samples taken from more/less informed trading subsets of races (fitted normal curves shown).

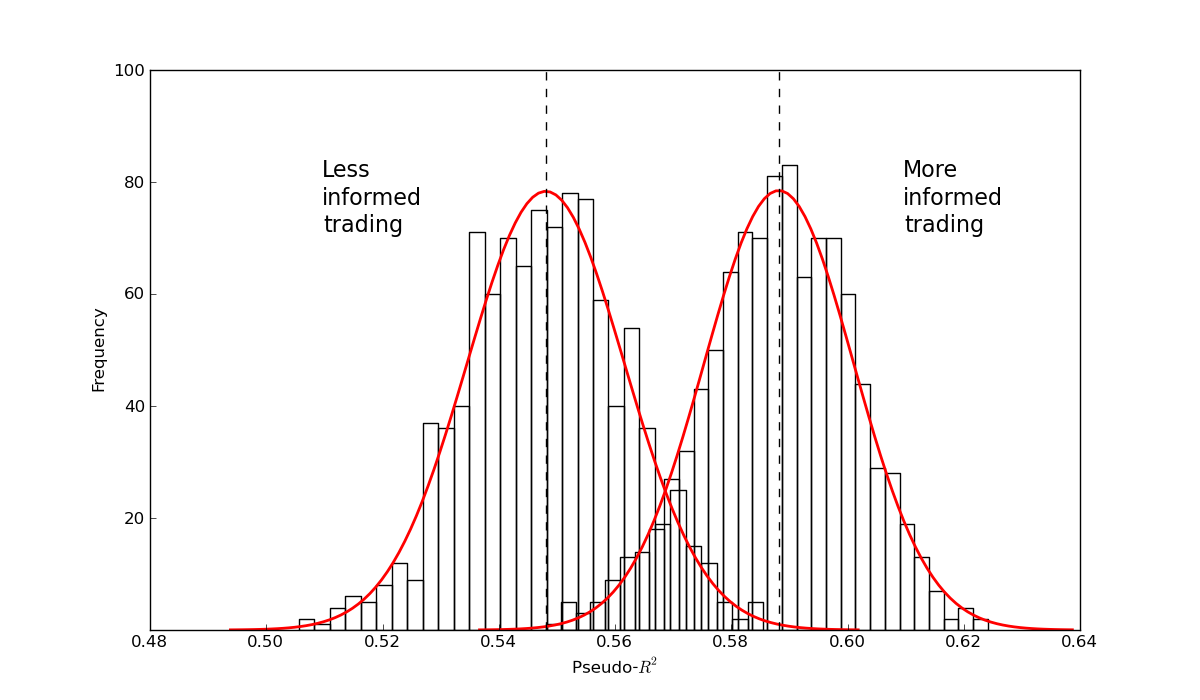
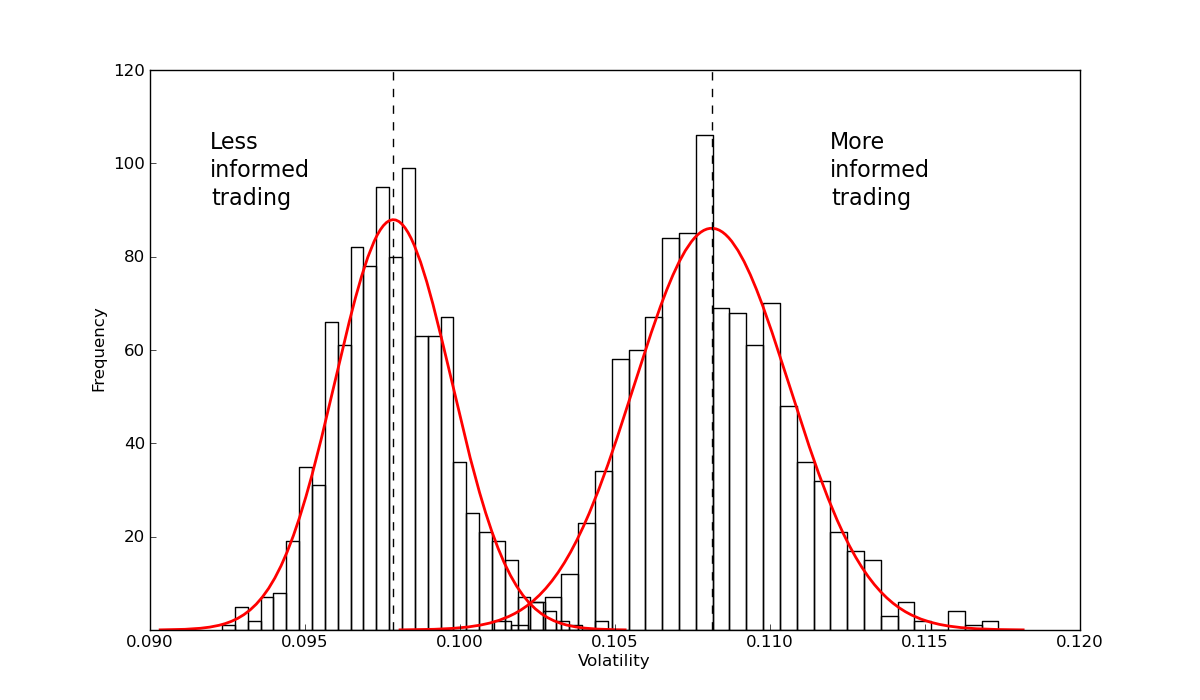


Figure 3. Histograms of volatility from 1000-bootstrap samples taken from more/less informed trading subsets of races (fitted normal curves shown).



**4. Conclusion**

We show that informed trading is associated with increased efficiency and volatility. Our findings are consistent with predictions that informed trading leads to more efficient outcomes, and that noise trading destabilizes asset values away from fundamental values, perhaps by limiting arbitrage of informed traders. Moreover, informed trading brings greater volatility, despite some predictions that noise trading is responsible for volatility. We shed light on the potential value and possible costs that traders with privileged information can bring to a financial market. For instance, focusing on innovative means of reducing liquidity risks to arbitrageurs, such as tighter controls on speculators and institutional noise traders, may help achieve efficient markets. By contrast, if calmer, less volatile markets are the goal, it is necessary to limit trading based on privileged information.

**References**

Bloomfield, R., M. O’Hara, and Saar, G. 2009. How noise trading affects markets: An experimental analysis. *Review of Financial Studies* 22:2275-302.

Campbell, J. Y., and A. S. Kyle. 1993. Smart money, noise trading and stock price behaviour. *The Review of Economic Studies* 60:1-34.

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann. 1990. Noise trader risk in financial markets. *Journal of Political Economy* 98:703-38.

French, K. R., and Roll, R. (1986). Stock return variances: the arrival of information and the reaction of traders. *Journal of Financial Economics*, 17(1), 5-26.

Gramm, M., and D. H. Owens. 2006. Efficiency in pari-mutuel betting markets across wagering pools in the simulcast era. *Southern Economic Journal* 72:926-37.

Johnson, J. E. V., O. Jones, and L. Tang. 2006. Exploring decision makers' use of price information in a speculative market. *Management Science* 52:897-908.

Johnstone, D. 2016. The effect of information on uncertainty and the cost of capital. *Contemporary Accounting Research* 32: 752-774.

Law, D., and D. A. Peel. 2002. Insider trading, herding behaviour and market plungers in the British horse-race betting market. *Economica* 69:327-38.

Maddala, G. S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, edited by P. Zarembka. New York: Academic Press, 105-42.

Peirson, J., and Smith, M. A. (2010). Expert analysis and insider information in horse race betting: regulating informed market behavior. *Southern Economic Journal*, 76(4), 976-992.

Sauer, R. D. 1998. The economics of wagering markets. *Journal of Economic Literature* 36:2021-64.

Shiller, R. J. 1990. Market volatility and investor behavior. *The American Economic Review* 80:58-62.

Shin, H. S. 1993. Measuring the incidence of insider trading in a market for state-contingent claims. *The Economic Journal* 103:1141-53.

Smith, M. A., D. Paton, and L. Vaughan Williams. 2006. Market efficiency in person-to-person betting. *Economica* 73:673-89.

Tavakoli, M., McMillan, D., and McKnight, P. J. 2012. Insider trading and stock prices. *International Review of Economics and Finance* 22:254-66.

Vaughan Williams, L., and D. Paton. 1997. Why is there a favorite-longshot bias in British racetrack betting markets? *The Economic Journal*, 107:150-8.