Impatient Trading, Liquidity Provision, and Stock Selection by Mutual Funds
– *Online Appendix*

Zhi Da, Pengjie Gao and Ravi Jagannathan

This Draft: January 25, 2010

This document contains supplementary materials to the paper titled "Impatient Trading, Liquidity Provision, and Stock Selection by Mutual Funds." It contains three parts. Appendix A provides a numerical illustration of our approach to decompose a mutual fund's stock selection skills. Appendix B provides a discussion of the variance decomposition procedure we used to examine the relative importance of different components of CS measures. Appendix C describes measures of the number of information events.

Appendix A: A Numerical Example for the Decomposition of Mutual Fund Stock Selection Skill

Assume there are six stocks (A, B, C, D, E, and F). A mutual fund's holdings in these stocks at the end of quarter $t-1$ (N_{t-1}) and t (N_t), stock prices at the end of quarter t (P_t), and the characteristics-adjusted stock returns during quarter $t+1$ [$R_{j,t+1} - BR_{t+1}(j,t)$] can be summarized in the following table:

Stock	N_{t-1}	N_t	P_t	$R_{j,t+1} - BR_{t+1}(j,t)$
A	2	1	10	−3%
B	2	0	15	−2%
C	2	2	20	−1%
D	2	2	25	1%
E	2	3	30	2%
F	0	2	35	3%

Stocks in the Hold, Buy and Sell categories are then defined by the holdings N_t^H, N_t^B and N_t^S:

Stock	$N_t^H = \min(N_{t-1}, N_t)$	$N_t^B = N_t - N_t^H$	$N_t^S = N_{t-1} - N_t^H$
A	1	0	1
B	0	0	2
C	2	0	0
D	2	0	0
E	2	1	0
F	0	2	0
Value	$H_t = 160$	$B_t = 100$	$S_t = 40$

The portfolio values H_t, B_t, and S_t are determined using the prices at the end of quarter t (P_t). Notice that $B_t > S_t$ – the difference is likely financed by fund inflows, a reduction in cash position or the sale of other non-stock assets held by the fund. The Hold, Buy, and Sell can be treated as three separate funds whose CS measures can be computed as:

	Hold	Buy	Sell
CS	$CS_{H,t+1} = 0.63\%$	$CS_{B,t+1} = 2.70\%$	$CS_{S,t+1} = -2.25\%$

Given this information, we then decompose the total CS measure into three components:

CS_{t+1}	CS_{t+1}^O	CS_{t+1}^T	CS_{t+1}^{adj}
1.42%	0.05%	1.49%	−0.12%

If we further assume that the fund trades B and F in the same direction as the aggregate order imbalance and trades A and E against the direction of aggregate order imbalance, we can further decomposes the trade component (CS_{t+1}^T) into an impatient trading component (C^{imp}) and a liquidity provision component (CS_{t+1}^{liq}):

CS_{t+1}^T	C^{imp}	CS_{t+1}^{liq}
1.49%	1.11%	0.38%

Appendix B: Variance Decomposition of the "Characteristic Selectivity" (CS) Measure

Empirically, we decompose the total "Characteristic Selectivity" (CS) measure (DGTW, 1997) into four components:[1]

$$CS = CS^O + CS^{adj} + C^{imp} + CS^{liq}.$$

Consequently, we have

$$var(CS) = cov(CS, CS^O) + cov(CS, CS^{adj}) + cov(CS, C^{imp}) + cov(CS, CS^{liq}),$$

where $var\,(\cdot)$ and $cov\,(\cdot)$ are the cross-sectional variance and covariance, respectively. Dividing both sides of the above equation by $var(CS)$, we then have

$$1 = \beta_P + \beta_{adj} + \beta_{inf} + \beta_{liq}.$$

The term $\beta_{(\cdot)}$ then measures the contribution of component $(\cdot)$ to the cross-sectional variation of CS. The sum of the contributions from the four components is equal to one by construction. β can be measured by regression. For instance, β_P is estimated by regressing CS^O on CS cross-sectionally. Empirically, we have a panel data of cross-sectionally demeaned CS, CS^O, CS^{adj}, C^{imp} and CS^{liq}. To estimate β, we run a Weighted Least Squares (WLS) regression. In practice, this means deflating the data for each fund-quarter by the number of funds in the corresponding cross-section.

Appendix C: Measures of Private Information Events — A Brief Description

Easley and O'Hara, along with their coauthors, in a series of papers develop this measure to capture the probability of information-based trading. Let α denote the probability that an information event occurs in a day; δ denote the low-state value of the underlying asset, conditional on the occurrence of an informational event; μ is the rate of informed trade arrivals; ϵ_b is the arrival rate of uninformed buy orders; and ϵ_s is the arrival rate of uninformed sell orders. Easley, Hvdkjaer and O'Hara (2002) propose the following MLE estimation to estimate the parameter

[1] For simiplicity of notation, we omit the time subscript t and fund superscript i.

vector $\Theta \equiv \{\alpha, \mu, \epsilon_b, \epsilon_s, \delta\}$

$$\begin{aligned} L(\Theta|B,S) &= (1-\alpha)\, e^{-\epsilon_b} \frac{\epsilon_b^B}{B!} e^{-\epsilon_s} \frac{\epsilon_s^S}{S!} \\ &\quad + \alpha\delta e^{-\epsilon_b} \frac{\epsilon_b^B}{B!} e^{-(\mu+\epsilon_s)} \frac{(\mu+\epsilon_s)^S}{S!} \\ &\quad + \alpha(1-\delta)\, e^{-(\mu+\epsilon_s)} \frac{(\mu+\epsilon_b)^B}{B!} e^{-\epsilon_s} \frac{\epsilon_s^S}{S!} \end{aligned} \tag{1}$$

where B and S represent total buy trades and sell trades for the day respectively. Given the above specifications, the probability of information-based trading, PIN, is

$$PIN = \frac{\alpha\mu}{\alpha\mu + \epsilon_b + \epsilon_s}. \tag{2}$$

With some independence assumptions across trading days, the likelihood function (1) becomes

$$L\left(\Theta|\,(B_i, S_i)_{i=1}^N\right) = \prod_{i=1}^{N} L(\Theta|B_i, S_i)\,. \tag{3}$$

The problem with estimation of the PIN measure is that in later years (since 2001), the number of buy and sell orders becomes extremely large, particularly for some NASDAQ stocks. One way to solve this problem is to impose the constraint that the arrival rates of informed and uninformed orders are the same,

$$\epsilon_b = \epsilon_s = \epsilon, \tag{4}$$

hence we estimate a modified version of (1),

$$L(\Theta|B,S) = (1-\alpha)\, e^{-2\epsilon} \frac{\epsilon^{B+S}}{B!S!} + \alpha\delta e^{-(\mu+2\epsilon)} \frac{\epsilon^B (\mu+\epsilon)^S}{B!S!} + \alpha(1-\delta)\, e^{-(\mu+2\epsilon)} \frac{\epsilon^S (\mu+\epsilon)^B}{B!S!} \tag{5}$$

and consequently, the probability of informed trading, PIN, is

$$PIN = \frac{\alpha\mu}{\alpha\mu + 2\epsilon}. \tag{6}$$

It is interesting to note that the probability that an information event occurs (α) and the rate of informed trade arrivals (μ) enter PIN as a product term ($\alpha\mu$). Although α and μ may be individually estimated rather imprecisely, since estimation errors in these two parameters are usually strongly negatively correlated, the resulting PIN estimate is quite precise. In addition, the variation in α and μ are offsetting, making PIN a much more stable measure bounded between 0 and 1.

Duarte and Young (2007) extend (1) to take into account large buy and sell volatilities, and pervasive positive correlation between buy and sell orders. Their model allows the possibility of order flow shocks and different distributions of the number of the buyer-initiated informed trades and seller-initiated informed trades. With such an extension, one may estimate an adjusted version of the probability of informed trading (AdjPIN) as

$$AdjPIN = \frac{\alpha \times [(1-\delta) \times \mu_b + \delta \times \mu_s]}{\alpha \times [(1-\delta) \times \mu_b + \delta \times \mu_s] + (\Delta_b + \Delta_s) \times \big[\alpha \times \theta' + (1-\alpha) \times \theta\big] + \epsilon_b + \epsilon_s} \tag{7}$$

where the additional parameter θ denotes the probability of symmetric order flow shocks conditional on no arrival of private information events, and θ' denotes the probability of symmetric order flow shocks conditional on the arrival of private information. Δ_b and Δ_s denote the additional arrival rate of buy orders and sell orders conditional on the arrival of the symmetric order flow shocks. Duarte and Young (2007) simplify (7) by restricting $\theta = \theta'$. To reduce the sheer volume of calculations, and to estimate a relatively parsimonious model with fewer parameters, we further impose the constraints that $\mu_b = \mu_s = \mu$ and $\Delta_b = \Delta_s = \Delta$. According to Duarte and Young (2007), the adjusted-PIN estimated with these constraints generate similar results to their full-fledged model.

Thus, the adjusted-PIN measure we estimate is specified as:

$$AdjPIN = \frac{\alpha \times \mu}{\alpha \times \mu + 2 \times \Delta \times \theta + 2 \times \epsilon}$$

In addition to causing large order imbalance, informed-trading will also force the market maker to increase the bid-ask spread. In the structural model of intra day trading costs proposed by Madhavan et. al. (1997), the price change can be captured by:

$$p_t - p_{t-1} = (\phi + \theta)x_t - (\phi + \rho\theta)x_{t-1} + u_t$$

Here x_t is the sign of the order flow (1: trade at ask, -1: trade at bid, 0: trade between bid and ask),

ϕ is the market maker's cost of supplying liquidity, ρ is the autocorrelation of the order flow, and θ captures the sensitivity of beliefs to unexpected order flows or the degree of private information. θ is therefore known as the information asymmetry component of the bid-ask spread and serves as an alternative measure of private information events. ϕ, ρ and θ will be jointly estimated with transaction level data using GMM on a quarterly basis.

To the extent that significant information events usually lead to abnormal trading in a stock, our last alternative measure is a measure of abnormal turnover (*aturn*) calculated in a similar fashion as in Chordia, Huh, and Subrahmanyam (2007). At the end of month t, for each stock, we estimate a regression in a 36-month rolling window $[t-35, t]$:

$$turn = a + bx + \varepsilon$$

where *turn* is monthly stock turnover defined as the ratio between total number of shares traded during the month and total number of shares outstanding, and x is a vector of adjustment regressors including 11 monthly dummy variables for months (January - November) as well as the linear and quadratic time-trend variables. The residual term for month t, ε_t, after standardization is the measure of abnormal turnover (*aturn*).

References

[1] Chordia, Tarun, Sahn-Wook Huh, and Avanidhar Subrahmanyam, 2007, The cross-section of expected trading activity, *Review of Financial Studies* 20, 709-740.

[2] Duarte, Jefferson, and Lance Young, 2009, Why is PIN priced? *Journal of Financial Economics*, Issue 2, 119-138.

[3] Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2002, Is information risk a determinant of asset returns?, *Journal of Finance* 57, 2185-2221.

[4] Madhavan, Ananth, Matthew Richardson and Mark Roomans, 1997, Why do security prices change? A transaction-level analysis of NYSE stocks, *Review of Financial Studies* 10, 1035-1064.