香港城市大學
City University of Hong Kong

Research Centre for International Economics

Working Paper: 2013048

Title of Paper

Estimation of Time Varying Adjusted Probability of Informed Trading and Probability of Symmetric Order-Flow Shock

Authors' List

Daniel Preve, City University of Hong Kong
Yiu-Kuen Tse, Singapore Management University

Abstract

Recently Duarte and Young (2009) study the probability of informed trading (PIN) proposed by Easley et al. (2002) and decompose it into two parts: the adjusted PIN (APIN) as a measure of asymmetric information and the probability of symmetric order-flow shock (PSOS) as a measure of illiquidity. They provide some cross-section estimates of these measures using daily data over annual periods. In this paper we propose a method to estimate daily APIN and PSOS by extending the method in Tay et al. (2009) using high-frequency transaction data. Our empirical results show that while PIN is positively contemporaneously correlated with variance, APIN is not. On the other hand, PSOS is positively correlated with daily average effective spread and variance, which is consistent with the interpretation of PSOS as a measure of illiquidity. Compared to APIN, PSOS exhibits clustering and sporadic bursts over time.

ESTIMATION OF TIME VARYING ADJUSTED PROBABILITY OF INFORMED TRADING AND PROBABILITY OF SYMMETRIC ORDER-FLOW SHOCK

DANIEL PREVE† AND YIU-KUEN TSE‡

JUNE 2012

ABSTRACT. Recently Duarte and Young (2009) study the probability of informed trading (PIN) proposed by Easley *et al.* (2002) and decompose it into two parts: the adjusted PIN (APIN) as a measure of asymmetric information and the probability of symmetric order-flow shock (PSOS) as a measure of illiquidity. They provide some cross-section estimates of these measures using daily data over annual periods. In this paper we propose a method to estimate daily APIN and PSOS by extending the method in Tay *et al.* (2009) using high-frequency transaction data. Our empirical results show that while PIN is positively contemporaneously correlated with variance, APIN is not. On the other hand, PSOS is positively correlated with daily average effective spread and variance, which is consistent with the interpretation of PSOS as a measure of illiquidity. Compared to APIN, PSOS exhibits clustering and sporadic bursts over time.

Key words and phrases. autoregressive conditional duration, market microstructure, probability of informed trading, probability of symmetric order-flow shock, transaction data.

† City University of Hong Kong. ‡ Singapore Management University. Address correspondence to Yiu-Kuen Tse, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore; e-mail: yktse@smu.edu.sg. We are grateful to conference participants at the 4th International Conference on Computational and Financial Econometrics (2010, London), SMU-ESSEC Symposium on Empirical Finance & Financial Econometrics (2011, Singapore) and seminar participants at Uppsala University and Helsinki Center of Economic Research for their comments and suggestions. All remaining errors are our own. The authors gratefully acknowledge research support from the Singapore Ministry of Education AcRF Tier 2 fund, research grant T206B4301-RS. The first author is thankful to the Sim Kee Boon Institute for Financial Economics, and the Institute's Centre for Financial Econometrics, at Singapore Management University for partial research support. Tao Yang provided excellent research assistance. Comments and suggestions from Tim Bollerslev, the Editor, and three anonymous referees are gratefully acknowledged.

1. Introduction

Since the seminal work of Easley *et al.* (1996), Easley *et al.* (1997) and Easley *et al.* (2002, EHO), many studies in the finance literature have used the probability of informed trading (PIN) to analyze the effects of asymmetric information on asset pricing and volatility. Easley and O'Hara (2004) argued that the effect of asymmetric information is undiversifiable and is thus priced. Hence, as a proxy for information asymmetry, PIN is expected to be significantly positively correlated with average stock returns.

Recently Duarte and Young (2009, DY) extended the EHO framework to analyze PIN as a measure of asymmetric information. Apart from relaxing the assumption that the arrival rate of informed sellers is the same as the arrival rate of informed buyers, as was imposed by EHO, they introduce what they call a "symmetric order-flow shock" to the model. They argue that traders may disagree on the interpretation of a public news event, which may cause both buy- and sell-orders to arrive at higher rates. As a result, DY propose a modification of PIN to measure the probability of informed trading, called the adjusted PIN (APIN). More importantly, they introduce a measure called the probability of symmetric order-flow shock (PSOS), which is the unconditional probability that a given trade comes from a shock to both buy- and sell-order flows. The authors show that high PSOS firms are usually firms with low trading volumes on most days, but who experience large increases in both buy- and sell-orders on days with public news. To this extent, they argue that PSOS is effectively a proxy for illiquidity, which is supported by their empirical finding that high PSOS firms tend to have high Amihud (2002) measures. Furthermore, they find that APIN is not priced, while PSOS is priced.

The empirical results of DY are based on the analysis of daily stock data over annual subperiods, for which the parameters of their APIN model are assumed to be constant within each year. Indeed, the empirical literature on PIN typically assumes constant probabilities of news and buy-sell intensity parameters over the sample period. The commonly adopted methodology is to estimate PIN using daily aggregates of buy- and sell-orders, which are assumed to be statistically independent. In addition to the assumption of constant probabilities of no news, good news and bad news, trade volume is not taken into account. These limitations have been criticized recently as possible causes for the anomalous behavior of PIN in some studies (see, e.g., Aktas *et al.*, 2007). To overcome these difficulties Tay *et al.* (2009, TTTW) consider the estimation of PIN using transaction data. Their model allows the probabilities of the states of news to vary daily, and they incorporate the use of covariates such as volume and duration of trade for the determination of PIN. An application of the TTTW model can be found in Hui *et al.* (2011).

In this paper we consider the estimation of APIN and PSOS using high-frequency transaction data by extending the methodology of TTTW. Following TTTW, we model transaction duration using the asymmetric autoregressive conditional duration (AACD) model proposed by Bauwens and Giot (2003). We allow the expected duration of buy- and sell-orders to be dependent on covariates such as lagged duration, lagged conditional expected duration, lagged trade direction (buy- or sell-order) and lagged trade volume. Also, we incorporate into our model time varying probabilities of no news, good news, bad news and common shocks, as featured in the DY model. The model parameters are estimated using the maximum likelihood estimation (MLE) method, from which we obtain daily estimates of APIN and PSOS. Our results provide an enhanced methodology to study the effects of asymmetric information and illiquidity on asset pricing. While the DY methodology provides a means for cross-sectional analysis of the relation of APIN and PSOS with size, spread and other illiquidity measures of a sample of stocks, our daily APIN and PSOS estimates can be used to trace the time-varying relation between asymmetric information, illiquidity and the price dynamics of each stock.

In sum, this paper contributes to the PIN literature in two aspects. First, while TTTW considered only events of good news, no news and bad news, we extend their method to incorporate the possibility of symmetric order-flow shock, which is achieved by modifying the conditional expected duration function according to different events. Thus, we are able to segregate PIN into the APIN and PSOS components. Second, by modeling time varying probability we estimate PIN, APIN and PSOS daily, which extends the cross-sectional analysis of DY to a dynamic context.

Our results on the contemporaneous correlations between PIN/APIN/PSOS and spread/return/variance are as follows. First, PSOS is significantly positively correlated with spread and variance, confirming its role as a measure of illiquidity. Second, the correlations between PIN and APIN with return are ambiguous as these two measures are not directional. On the other hand, return is positively (negatively) correlated with the probability of good (bad) news, due to the directional nature of the latter. Third, while PIN is positively correlated with variance, the correlations between APIN and variance is ambiguous.

The remainder of the paper is organized as follows. In Section 2 we briefly review the PIN model of EHO and the APIN model of DY. In Section 3 we review the PIN-AACD model of TTTW and outline our extension, the APIN-AACD model. In doing so, we also briefly review the AACD model of Bauwens and Giot (2003). Section 4 reports the results of our empirical study. Section 5 concludes.

2. The PIN and APIN Models

In this section we briefly summarize the PIN model of EHO and its extension, namely, the APIN model of DY. A more extensive review can be found in the supplementary web-based appendix.

2.1. **The PIN Model.** Let B_d and S_d denote the aggregate number of buy- and sell-orders on day d, respectively. In the PIN model, B_d and S_d are assumed to be independent Poisson random variables, with different intensities for days with bad news (B), good news (G) and no news (N). Let θ_E denote the probability of news being released on day d and let θ_B denote the probability of bad news, conditional on the release of news. Thus, the daily state probabilities are $\pi_B = \theta_E\theta_B$, $\pi_G = \theta_E(1-\theta_B)$ and $\pi_N = 1-\theta_E$, for a day with bad news, good news and no news, respectively. For a day with no news, the means of B_d and S_d are λ_1 and λ_{-1}, respectively. For a day with bad news the sell intensity increases by a constant δ, while the buy intensity remains the same as for a day with no news. Similarly, for a day with good news the buy intensity increases by δ, while the sell intensity stays the same as for a no-news day. EHO compute the PIN as the relative intensity of informed trades to the intensity of all trades, so that

$$\mathrm{PIN} = \frac{\mathrm{P}_2}{\mathrm{P}_1 + \mathrm{P}_2}, \tag{2.1}$$

where $\mathrm{P}_1 = \lambda_1 + \lambda_{-1}$ and $\mathrm{P}_2 = \theta_E \delta$.

2.2. **The APIN Model.** DY extended the PIN model of EHO by allowing for the arrival rate of informed sellers (δ_{-1}) to be different from the arrival rate of informed buyers (δ_1) and, more importantly, by allowing both buy- and sell-order flows to increase on certain days even when there is no news. In the APIN model B_d and S_d have different intensities for days with bad news and a common shock (CB), good news and a common shock (CG) and no news and a common shock (CN). Let θ_C denote the daily probability of a common shock.[1] In the event of a common shock, the buy intensity increases by $\triangle_1$ and the sell intensity by $\triangle_{-1}$. DY compute the APIN as

$$\mathrm{APIN} = \frac{\mathrm{P}_2}{\mathrm{P}_1 + \mathrm{P}_2 + \mathrm{P}_3}, \tag{2.2}$$

and introduce PSOS as

$$\mathrm{PSOS} = \frac{\mathrm{P}_3}{\mathrm{P}_1 + \mathrm{P}_2 + \mathrm{P}_3}, \tag{2.3}$$

where $\mathrm{P}_1 = \lambda_1 + \lambda_{-1}$, $\mathrm{P}_2 = \theta_E[(1-\theta_B)\delta_1 + \theta_B\delta_{-1}]$ and $\mathrm{P}_3 = \theta_C(\triangle_1 + \triangle_{-1})$. Note that APIN in Equation (2.2) reduces to the original PIN measure in Equation (2.1) when $\delta_1 = \delta_{-1}$ and $\theta_C = 0$, as expected.

3. The PIN-AACD and APIN-AACD Models

In this section we review the PIN-AACD model of TTTW and outline our extension, the APIN-AACD model, analogous to the extension by DY of the PIN model of EHO. In doing so, we first review the AACD model of Bauwens and Giot (2003).

3.1. **The AACD Model of Trade Direction.** TTTW model trade direction (buy- and sell-initiated order) and duration between trades (waiting time) jointly using an AACD model, and compute PIN from this model. We denote x_i as the (diurnally adjusted) waiting time between trade $i-1$ at time t_{i-1} and trade i at time t_i so that $x_i = t_i - t_{i-1}$. In addition, we denote y_i as the trade direction of the ith trade, which takes on values -1 and 1 representing a sell- and buy-order, respectively. Φ_{i-1} denotes the information upon the $(i-1)$th trade, which may include trade direction y_{i-1}, transaction volume v_{i-1}, waiting time x_{i-1}, as well as their lagged values.

Conditional on Φ_{i-1}, TTTW assume that both *potential* trade directions (buy or sell) at time t_i follow latent point processes. More specifically, given Φ_{i-1}, $\{B_i(s_i),\, s_i \geq 0\}$ and $\{S_i(s_i),\, s_i \geq 0\}$ are latent Poisson processes, representing buy- and sell-orders, with common start time t_{i-1}, i.e. $s_i = s_i(t) = t - t_{i-1}$, and intensities λ_{1i} and $\lambda_{-1,i}$. The observed trade direction at time t_i is the outcome of the competition between the two Poisson processes to be the *first* arrival.

[1]DY also consider models for which the probability of common shock varies with the state of news or no news. However, they argue empirically that the restriction of imposing invariance is innocuous. In this paper we adopt this restriction.

Conditional on Φ_{i-1}, let d_{ji} be the latent waiting time (duration) until the first occurrence for trade direction j and suppose that d_{1i} is independent of $d_{-1,i}$. Let $x_i = \min\{d_{1i}, d_{-1,i}\}$ and $y_i = j$, where $j = -1$ if $d_{-1,i} = x_i$ and $j = 1$ if $d_{1i} = x_i$. Under the Poisson process assumption d_{ji} is exponentially distributed with mean $\psi_{ji} = 1/\lambda_{ji}$ given Φ_{i-1}. It can be shown that x_i is conditionally exponential with mean $1/(\lambda_{1i} + \lambda_{-1,i})$, and y_i is conditionally two-point distributed with probability function $\lambda_{ji}/(\lambda_{1i} + \lambda_{-1,i})$, for $j = -1, 1$. Moreover, it can be shown that x_i and y_i are independent conditional on Φ_{i-1}. Hence, the conditional joint distribution (pf-pdf) of duration x_i and direction y_i is given by

$$f(x_i, y_i|\Phi_{i-1}) = \lambda_{ji} e^{-(\lambda_{1i}+\lambda_{-1,i})x_i} \tag{3.1}$$
$$= \left(\frac{1}{\psi_{1i}}\right)^{\mathbf{1}_{\{y_i=1\}}} \left(\frac{1}{\psi_{-1,i}}\right)^{\mathbf{1}_{\{y_i=-1\}}} \exp\left[-\left(\frac{1}{\psi_{1i}} + \frac{1}{\psi_{-1,i}}\right) x_i\right],$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function.

While we adopt the exponential waiting time assumption as a consequence of the Poisson assumption of the arrival of trade orders in the EHO framework, alternative waiting time assumptions can be considered. For example, if Weibull distributions with identical shape parameters, which encompass the exponential, are used for the latent waiting time variables the conditional independence of x_i and y_i still holds (see Bauwens and Giot, 2003).[2] From an empirical perspective, Bauwens *et al.* (2004) compared the predictive performance of various duration distributions and concluded that none are clearly preferred over the exponential.[3] Tay *et al.* (2011) estimated the AACD model using both exponential and Weibull assumptions, and the results were found to be similar. Tse and Yang (2012) fitted ACD models using the exponential assumption and semiparametric method. They reported similar results when these estimates are used to calculate intraday volatility. Thus, the exponential assumption can be viewed as a theoretical consequence of the EHO model with some support from the empirical literature.

3.2. **The PIN-AACD Model.** Like the PIN model of EHO, the PIN-AACD model of TTTW has states corresponding to no news, good news and bad news. However, unlike the EHO approach, TTTW allow for interactions between consecutive buy- and sell-orders, and account for the duration between trades and the volume of the trade. Note that the PIN-AACD model allows PIN to be computed for a specific day as well as over intraday intervals using high-frequency transaction data.

In the PIN-AACD model, the conditional expected duration ψ^s_{ji} of d^s_{ji} for $s \in \mathcal{S} = \{G, B, N\}$ is based on ψ^N_{ji} (the conditional expected duration on a no-news day), where ψ^N_{ji} is assumed to follow the extended logarithmic ACD(1,1) model

$$\ln \psi^N_{ji} = \nu_{j1}\mathbf{1}_{\{y_{i-1}=1\}} + \nu_{j,-1}\mathbf{1}_{\{y_{i-1}=-1\}} + \beta_j \ln \psi^N_{j,i-1} + \alpha_j \ln x_{i-1} + \varsigma_j y_{i-1} \ln v_{i-1}, \tag{3.2}$$

[2]As shown by Drost and Werker (2004) the exponential assumption has an advantage that the ML estimator of the conditional expected duration is consistent provided the conditional expected duration equation is correctly specified, regardless of the duration distribution. It should be noted, however, that Drost and Werker's (2004) result applies to linear ACD models and is not applicable to nonlinear models such as the AACD model.

[3]Bauwens *et al.* (2004) pointed out that financial duration models need to put more probability mass on small durations, and as far as this is concerned, the generalized gamma distribution offers an improvement.

for $j = -1, 1$, where v_{i-1} is the volume of the trade at time t_{i-1}.[4] Thus, the base equation for ψ^N_{ji} depends on whether the previous transaction is a buy- or sell-initiated order, y_{i-1}, the lagged conditional expected duration, $\psi^N_{j,i-1}$, the previous duration, x_{i-1}, and the lagged signed logarithmic volume, $y_{i-1} \ln v_{i-1}$. Hence, in contrast to the PIN model of EHO, the PIN-AACD model allows volume to impact trade intensity. Analogous to the PIN model of EHO, on a bad-news day $\ln \psi^B_{-1,i} = \ln \psi^N_{-1,i} - \mu_B$ and on a good-news day $\ln \psi^G_{1i} = \ln \psi^N_{1i} - \mu_G$.[5] The equations for the implied conditional Poisson intensities, λ^s_{ji}, are $\lambda^G_{1i} = \lambda^N_{1i}(e^{\mu_G} - 1)$ and $\lambda^B_{-1,i} = \lambda^N_{-1,i}(e^{\mu_B} - 1)$, where $\lambda^N_{ji} = 1/\psi^N_{ji}$ with ψ^N_{ji} as in Equation (3.2). We expect the parameters μ_B and μ_G to be positive so that on a bad-news (good-news) day, $\psi^B_{-1,i}$ (ψ^G_{1i}) decreases and $\lambda^B_{-1,i}$ (λ^G_{1i}) increases.

In their general specification, TTTW model θ_E and θ_B rather than assuming them to be constant, thus allowing the probabilities of news to vary over time. More specifically, they assume logistic models in which the arrival of bad news, good news and no news on day d depends on the aggregate volume of buy- and sell-orders. This is motivated by recent empirical work reporting positive correlation between (public) information and trading volume.[6] TTTW denote the average number of lots traded per day initiated by buy orders by $\overline{V}^B$. Similarly, they denote the average number of lots traded per day initiated by sell orders by $\overline{V}^S$. The numbers of lots traded on day d initiated by buy- and sell-orders are denoted by V^B_d and V^S_d, respectively. The probability of news on day d is assumed to be

$$\theta_{Ed} = 1 - \frac{1}{1 + \exp\left\{\gamma_1 + \gamma_2\left[\ln\left(V^B_d + V^S_d\right) - \ln\left(\overline{V}^B + \overline{V}^S\right)\right]\right\}},$$

where γ_2 is expected to be strictly positive.[7] Given news on day d, the probability of bad news is assumed to be

$$\theta_{Bd} = 1 - \frac{1}{1 + \exp\left[\gamma_3\left(\ln V^S_d - \ln \overline{V}^S\right) - \gamma_4\left(\ln V^B_d - \ln \overline{V}^B\right)\right]},$$

where γ_3 and γ_4 are expected to be strictly positive. The arrival of bad news, good news and no news on day d are given by $\pi_{Bd} = \theta_{Ed}\theta_{Bd}$, $\pi_{Gd} = \theta_{Ed}(1 - \theta_{Bd})$ and $\pi_{Nd} = 1 - \theta_{Ed}$, respectively.

[4] TTTW use the log-ACD(1, 1) model of Bauwens and Giot (2000) as a basis for (3.2), rather than the standard ACD(1,1) model of Engle and Russell (1998), as it is flexible for including additional explanatory variables in the autoregressive equation.

[5] In fact, TTTW assume that $\mu_B = \mu_G$. Here we allow μ_B to be different from μ_G, as might be justified due to short-selling restrictions (see Diamond and Verrecchia, 1991).

[6] Andersen (1996) proposed the modified mixture of distribution hypothesis in which volatility and volume of informed traders are both driven by information intensity. This model is further extended by Li and Wu (2006). Berry and Howe (1994) found evidence in support of trading volume driven by information as measured by Reuters' news release, with similar findings reported by Mitchell and Mulherin (1994). To circumvent the difficulty in identifying "relevant information", we follow TTTW to adopt a reduced-form modeling approach for information intensity with trading volume as a proxy. To further distinguish between asymmetric information and differential information, the approach proposed by Sarkar and Schwartz (2009) in modeling "market sidedness" may be considered.

[7] The state-probability functions proposed in this paper assume the volume series are stationary, which may not be valid over a long period of time. For the data used in this paper, however, we do not observe trends in the volume series that would suggest the violation of a stable mean level.

The PIN-AACD model is estimated using the MLE method. With $N_d = B_d + S_d$ orders on day d, its likelihood function is given by

$$\prod_{d=1}^{D}\left[\sum_{s\in\mathcal{S}}\pi_{sd}\left(\prod_{i=1}^{N_d} f_s(x_i, y_i|\Phi_{i-1})\right)\right], \qquad \text{for } \mathcal{S} = \{B, G, N\}, \tag{3.3}$$

where

$$f_s(x_i, y_i|\Phi_{i-1}) = \left(\frac{1}{\psi_{1i}^s}\right)^{\mathbf{1}_{\{y_i=1\}}}\left(\frac{1}{\psi_{-1,i}^s}\right)^{\mathbf{1}_{\{y_i=-1\}}} \exp\left[-\left(\frac{1}{\psi_{1i}^s}+\frac{1}{\psi_{-1,i}^s}\right)x_i\right],$$

whenever $x_i \geq 0$, $y_i = -1, 1$ and zero otherwise (cf. Equation 3.1).

Because of the Poisson process assumption, it can be shown that conditional on Φ_{i-1} the expected number of trades due to all traders in the *fixed* interval $(t_{i-1}, t_i]$ on day d is

$$\mathrm{E}[B_i(x_i) + S_i(x_i)|\Phi_{i-1}, x_i] = \sum_{s\in\mathcal{S}}\pi_s \mathrm{E}[B_i(x_i) + S_i(x_i)|\Phi_{i-1}, x_i, s] = \big(\underbrace{\lambda_{1i}^N + \lambda_{-1,i}^N}_{\mathrm{P}_{1i}} + \underbrace{\pi_{Gd}\lambda_{1i}^G + \pi_{Bd}\lambda_{-1,i}^B}_{\mathrm{P}_{2i}}\big)x_i,$$

where P_{1i} and P_{2i} are due to uninformed and informed trades, respectively.[8] Similar to the original PIN measure of EHO, TTTW compute the daily AACD PIN as

$$\mathrm{PIN}_d = \frac{\sum_{i=1}^{N_d}\mathrm{P}_{2i}x_i}{\sum_{i=1}^{N_d}(\mathrm{P}_{1i}+\mathrm{P}_{2i})x_i}, \tag{3.4}$$

emphasizing that Equation (3.4) can be modified to compute AACD PIN measures over intraday intervals, in which case the summations are over trades in specific intraday intervals.

3.3. **The APIN-AACD Model.** Like the APIN model of DY, our proposed APIN-AACD model has six different daily states, allowing for symmetric order-flow shock trades. Similar to the PIN-AACD model, and in contrast to the APIN model, the APIN-AACD model allows for the APIN and the PSOS to be computed daily as well as over intraday intervals.

As an extension of the PIN-AACD model, the APIN-AACD model has three additional states representing trading days in which the conditional intensities of both B_d and S_d increase due to common shocks. These days occur with probabilities θ_{Cd}. Analogous to TTTW, we assume a logistic model for the daily probability of a common shock such that

$$\theta_{Cd} = 1 - \frac{1}{\exp\left[\gamma_5\left(\ln V_d^B - \ln \overline{V}^B\right)_+\left(\ln V_d^S - \ln \overline{V}^S\right)_+\right]},$$

where $(u)_+$ equals u if $u > 0$ and zero otherwise. Note that, for θ_{Cd} to lie between 0 and 1, we must have $\gamma_5 \geq 0$. Furthermore, $\theta_{Cd} = 0$ unless $V_d^B > \overline{V}_d^B$ and $V_d^S > \overline{V}_d^S$. Thus, there is no symmetric order-flow shock on day d unless both the buy- and sell-orders on that day are larger than their corresponding sample average. This assumption appears to be reasonable given that a symmetric order-flow shock induces both buy and sell orders. In practice θ_{Cd} is frequently zero. For example, for the IBM data in our empirical study 515 out of 754 probabilities (i.e., θ_{Cd}) are zero.

[8]By definition, $\lambda_{1i}^G = 1/\psi_{1i}^G - \lambda_{1i}^N$ and $\lambda_{-1,i}^B = 1/\psi_{-1,i}^B - \lambda_{-1,i}^N$.

Like the PIN-AACD model, the conditional expected duration ψ^s_{ji} for each state $s \in \mathcal{S}$ is based on ψ^N_{ji} in Equation (3.2). Equations for the remaining ψ^s_{ji} are given in Table 1 (see the web-based appendix for an illustration). Analogous to the APIN model of DY, on a bad-news day with a common shock $\ln\psi^{CB}_{1i} = \ln\psi^N_{1i} - \mu_{CB}$ and $\ln\psi^{CB}_{-1,i} = \ln\psi^N_{-1,i} - \mu_B - \mu_{CB}$. Similarly, on a good-news day with a common shock $\ln\psi^{CG}_{1i} = \ln\psi^N_{1i} - \mu_G - \mu_{CG}$ and $\ln\psi^{CG}_{-1,i} = \ln\psi^N_{-1,i} - \mu_{CG}$. Finally, on a no-news day with a common shock $\ln\psi^{CN}_{ji} = \ln\psi^N_{ji} - \mu_{CN}$.[9]

It can be shown that conditional on Φ_{i-1} the expected number of trades due to all traders in the fixed interval $(t_{i-1}, t_i]$ on day d is

$$\begin{aligned}&\mathrm{E}[B_i(x_i) + S_i(x_i)|\Phi_{i-1}, x_i]\\ &= \big[\underbrace{\lambda^N_{1i} + \lambda^N_{-1,i}}_{\mathrm{P}_{1i}} + \underbrace{\pi_{Gd}\lambda^G_{1i} + \pi_{Bd}\lambda^B_{-1,i}}_{\mathrm{P}_{2i}} + \underbrace{\pi_{CBd}\big(\lambda^{CB}_{1i} + \lambda^{CB}_{-1,i}\big) + \pi_{CGd}\big(\lambda^{CG}_{1i} + \lambda^{CG}_{-1,i}\big) + \pi_{CNd}\big(\lambda^{CN}_{1i} + \lambda^{CN}_{-1,i}\big)}_{\mathrm{P}_{3i}}\big]x_i,\end{aligned}$$

where P_{1i}, P_{2i} and P_{3i} are the expected numbers of trades due to uninformed, informed and symmetric order-flow shock trades, respectively.[10] The daily AACD APIN and PSOS are given by

$$\mathrm{APIN}_d = \frac{\sum_{i=1}^{N_d} \mathrm{P}_{2i}x_i}{\sum_{i=1}^{N_d}(\mathrm{P}_{1i} + \mathrm{P}_{2i} + \mathrm{P}_{3i})x_i}, \tag{3.5}$$

and

$$\mathrm{PSOS}_d = \frac{\sum_{i=1}^{N_d} \mathrm{P}_{3i}x_i}{\sum_{i=1}^{N_d}(\mathrm{P}_{1i} + \mathrm{P}_{2i} + \mathrm{P}_{3i})x_i}. \tag{3.6}$$

Note that the APIN-AACD measure in Equation (3.5) reduces to the PIN-AACD measure in Equation (3.4) when $\theta_{Cd} = 0$, as expected.

4. Empirical Results

4.1. **Data.** The intraday data used in this section were extracted and compiled from the New York Stock Exchange (NYSE) Trade and Quote (TAQ) Database provided through the Wharton Research Data Services. We retrieved data from the Consolidated Trade (CT) file as well as the Consolidated Quote (CQ) file. From the CT file we downloaded the data for the date, trading time, price and number of shares traded for each stock in our study. From the CQ file we downloaded the data for the offer and bid prices, as well as the time of the quote revisions. The data sets used consist of high-frequency transaction data for the IBM, GE (General Electric), PG (Procter and Gamble) and WMT (Walmart) stocks over the period Jan 1, 2005 through Dec 31, 2007, covering 754 trading days.

Due to opening effects, the first 20 minutes (9:30 am to 9:50 am) of each trading day were removed. All transactions after 4:00 pm were also deleted. Days where the opening transaction occurred after the

[9]We also experimented with a more parsimonious model specification ($\mu_B = \mu_G$ and $\mu_{CB} = \mu_{CG}$) which yielded empirical results similar to those reported in Section 4.

[10]Analogous to TTTW, we define $\lambda^G_{1i} = 1/\psi^G_{1i} - \lambda^N_{1i}$ and $\lambda^B_{-1,i} = 1/\psi^B_{-1,i} - \lambda^N_{-1,i}$. In addition, we define $\lambda^{CB}_{1i} = 1/\psi^{CB}_{1i} - \lambda^N_{1i}$, $\lambda^{CB}_{-1,i} = 1/\psi^{CB}_{-1,i} - \lambda^N_{-1,i} - \lambda^B_{-1,i}$, $\lambda^{CG}_{1i} = 1/\psi^{CG}_{1i} - \lambda^N_{1i} - \lambda^G_{1i}$, $\lambda^{CG}_{-1,i} = 1/\psi^{CG}_{-1,i} - \lambda^N_{-1,i}$, $\lambda^{CN}_{1i} = 1/\psi^{CN}_{1i} - \lambda^N_{1i}$ and $\lambda^{CN}_{-1,i} = 1/\psi^{CN}_{-1,i} - \lambda^N_{-1,i}$.

first 20 minutes of the trading day or where there were insufficient (less than 10) transactions between 9:50 am and 10:00 am to obtain meaningful initial values for the ML estimation were also removed.

The frequency of zero trade durations (simultaneous transactions) in the data sets is high. For example, about 35% of the observations for the IBM data are of zero durations. We deal with the zero durations in the following way. For transactions with the same time stamp we aggregate the transaction volumes and compute an average price weighted by volume, as described in Pacurar (2008).

We compute the diurnal factors, which are linked to the trading habits and intraday seasonality, by applying a smoothing cubic spline to the average raw duration at each time point with available data. We use the Matlab function `spap2` to estimate the spline by least-squares. The cubic spline is made up of 6 polynomial pieces, with knots set on each hour (10:00 am to 4:00 pm). Following Engle and Russell (1998), we set the mean of the computed diurnal factors equal to the sample mean of the raw durations. Note that, in practice, this implies that the sample mean of our diurnally adjusted (DA) durations is approximately 1.

Like DY, we classify trade direction according to the Lee and Ready (1991) algorithm. Trades for which the algorithm does not apply were further classified as buyer- or seller-initiated based on a tick test. Some summary statistics of the resulting data sets are given in Table 2. The average number of trades per day ranges from 4,468.32 (PG) to 5,419.49 (GE). More than 50% of the trades for all stocks are sell orders. Table 2 also reports estimates of the state probabilities as well as PIN (APIN and PSOS) obtained using the EHO (DY) models. It can be observed that $\hat{\pi}_N$ of the EHO model exceeds $\hat{\pi}_N + \hat{\pi}_{CN}$ of the DY model for all stocks. While there is reduction in the probability of no news (with or without common shocks) of the DY model, the probability of good and bad news (with or without common shocks) correspondingly increases. The results also show that PSOS is larger than APIN for all stocks.

4.2. **Maximum Likelihood Estimation of the Models.** ML estimation of the PIN- and APIN-AACD models with time varying probabilities was performed using the Matlab function `fmincon` with the interior-point algorithm and numerical derivatives. The values ψ_{j1}^N used to initialize each day were computed as follows. Let n_d denote the number of transactions between 9:50 am and 10:00 am on trading day d. As initial values for day d we use

$$\psi_{j1}^N = \frac{\sum_{i=1}^{n_d} \mathbf{1}_{\{y_i=j\}} x_i}{\sum_{i=1}^{n_d} \mathbf{1}_{\{y_i=j\}}}, \qquad \text{for } j = -1, 1.$$

To search for a global optimum, we use a random starting point for the numerical method and run the likelihood optimization 10 times for each data set. We then select the maximum of these 10 optimizations. The estimation procedure converges for all data sets.

The ML estimation results for the PIN- and APIN-AACD models with time varying probabilities are presented in Table 3. It can be seen that the parameter estimates exhibit a remarkable resemblance across the four stocks. We note that $\hat{\gamma}_1$ is negative for all stocks, implying that the estimated probability of news $\hat{\theta}_{Ed}$ is less than 0.5 on an *average* day (when the buy- and sell-orders are equal to the sample average). As expected, estimates of γ_2 through γ_5 are all positive. Both $\hat{\beta}_1 + \hat{\alpha}_1$ and $\hat{\beta}_{-1} + \hat{\alpha}_{-1}$ are less than 1. The persistence of the latent processes, however, appears to be quite high. Similar to TTTW, we observe that $\hat{\varsigma}_{-1} > 0$ and $\hat{\varsigma}_1 < 0$ for both models and all stocks (although these estimates are not statistically significant for WMT in both models), implying that large buy orders induce shorter conditional expected durations for subsequent buy orders but longer conditional expected durations for sell orders. The opposite goes for large sell orders. Thus, the results suggest that volume plays an explicit part in predicting trade direction.[11]

Note that the standard regularity conditions for the limiting null distribution of the likelihood ratio test statistic to be chi-squared are not satisfied when testing the restricted ($\gamma_5 = 0$) PIN-AACD model against the unrestricted APIN-AACD model. This is because the parameter lies on the boundary of the parameter space under the null hypothesis. Consequently we do not report any likelihood ratio or Wald test results. However, both the Akaike Information Criterion (AIC) and the Schwarz Bayesian Criterion (SBC) support the selection of the unrestricted APIN-AACD model over the restricted PIN model.[12] Finally we note that all estimates are statistically significant at the 5 percent level, except for $\hat{\gamma}_2$ of PG for the APIN-AACD model as well as $\hat{\varsigma}_1$ and $\hat{\varsigma}_{-1}$ for WMT for both models.[13] The statistical insignificance of $\hat{\gamma}_2$ implies that the probability of news does not vary over different days. However, as $\hat{\gamma}_3$ and $\hat{\gamma}_4$ are statistically nonzero, the probabilities of good news and bad news are still time varying.

4.3. **Estimates of Daily PIN, APIN and PSOS.** Figure 1 presents the plots of the estimated daily probabilities of good news, no news and bad news for the PIN-AACD model applied to the IBM stock. It can be seen that the model-implied probability of bad news appears to be quite stable throughout the sample period and is less than 0.2 more than half of the days. In contrast, the estimated probability of good news is more volatile, with values exceeding 0.8 for a few days. Figure 2 shows the plots of the estimated daily state probabilities for the APIN-AACD model applied to the IBM stock. For this model, the estimated probability of good news without common shock ($\hat{\pi}_G$) is more stable over time compared to the probability of good news in the PIN-AACD model. In particular, $\hat{\pi}_G$ is less than 0.5 for all days. In contrast, estimates of the probabilities of events with common shock ($\hat{\pi}_{CG}$, $\hat{\pi}_{CN}$ and $\hat{\pi}_{CB}$) are irregular and sporadic. The estimated probabilities are zero for most days, but may be quite large (exceeding 0.5) on some days. This result suggests that the volatile pattern of $\hat{\pi}_G$ in the PIN-AACD model may be due

[11] While volume is statistically significant for three out of four stocks its economic effects are relatively small. The economic impact on conditional expected duration is the highest for lagged conditional expected duration, followed by lagged duration and then lagged signed volume.

[12] Comparing the results in Tables 2 and 3, we can see that the Ljung-Box statistics are reduced for the diurnally adjusted durations versus the raw durations and further drastically reduced for the standardized durations, although they are still highly significant. These results are in line with those in the literature (see, e.g., Engle and Russell (1998)), due to the enormous sample size of high-frequency data (note that though our Ljung-Box statistics are much larger than those of Engle and Russell (1998), our sample size is 78 to 88 times larger). However, the Ljung-Box statistics for the APIN-AACD model are not reduced versus the PIN-AACD model for two stocks. While there may be further improvement in the Ljung-Box statistics by considering higher order AACD models, this extension has not been considered in this paper.

[13] The standard errors are computed using the Hessian matrix. The inner product form of the asymptotic variance is not computable as the likelihood for each observation cannot be separated.

to common-shock trading. We also note that the average $\hat{\pi}_{CN}$ estimated from the APIN-AACD model (see the web-based appendix) is lower than $\hat{\pi}_{CN}$ (see Table 2) obtained from the DY model.

Daily AACD PIN, APIN and PSOS estimates were computed using Equations (3.4), (3.5) and (3.6), respectively. Figure 3 presents the plots of PIN/APIN/PSOS for the IBM stock.[14] APIN appears to be more stable than PIN. While APIN is less than 0.05 on almost all days, PIN fluctuates a lot with quite a few days exceeding 0.1. On the other hand, PSOS behaves quite differently from PIN and APIN. In particular, while PSOS is zero for many days, it also fluctuates to above 0.2 for quite a few days. We also note that PSOS may remain zero for an extended period of time, during which common-shock traders are absent from the market. Furthermore, the average PSOS and APIN computed from the APIN-AACD model (see the web-based appendix) are lower than the PSOS and APIN estimated using the DY model (see Table 2).

DY reported the correlations between PIN, APIN and PSOS computed over a period of time with some variables of interest such as spread and firm size across a cross-section of NYSE stocks. In Table 4 we present the contemporaneous correlations of the daily estimates of PIN, APIN and PSOS with the average daily effective spread, the daily variance and return for the four stocks.[15] In addition, we also present the contemporaneous correlations between daily return and the daily estimates of the probability of good news $\hat{\pi}_G$ and bad news $\hat{\pi}_B$ obtained from the APIN-AACD model. Our correlations computed using the time series data provide some complementary results to the cross-section analysis of DY.[16]

We observe that daily PSOS is significantly positively correlated with the average effective spread, with p-value ranging from 0.003 to 0.062. This result is consistent with the notion of PSOS being a measure of illiquidity, which is further supported by the positive correlation between PSOS and variance. In contrast, daily APIN is not significantly correlated with effective spread for all four stocks, and is indeed negatively significantly correlated with variance for three stocks. This rather surprising result raises doubts about the use of APIN as a measure of asymmetric information. It also raises the question of how information asymmetry may impact high-frequency volatility. On the other hand, PIN is significantly positively correlated with variance for all four stocks, while its correlation with spread is significant for only two stocks. Again, the positive correlation of variance with PIN but not APIN remains a puzzle.

Correlations between daily returns and PIN/APIN/PSOS are largely ambiguous. This result can be explained by the fact that the PIN, APIN and PSOS measures are not directional and do not determine the directions of price movements in the time-series context. However, the correlations between daily

[14]The plots for the other stocks of the PIN-AACD and APIN-AACD models are visually similar. Additional plots for the GE stock can be found in the web-based appendix.

[15]Effective spread is computed as two times the absolute value of price minus mid-quote. Daily variance is computed using the ACD-ICV method proposed by Tse and Yang (2012). This method estimates the integrated conditional variance (ICV) over an intraday interval using tick data. It is computed as the weighted sum of the instantaneous conditional variances estimated from an ACD model.

[16]DY also examined the determination of the expected stock returns by running the Fama-MacBeth regressions of time-series averages from firm-level cross sections. Our contemporaneous correlation analysis, however, does not have equilibrium asset pricing implications.

return and $\hat{\pi}_G$ ($\hat{\pi}_B$) are significantly positive (negative), which is consistent with the assumption that news moves the market.

Finally we observe that APIN is negatively contemporaneously correlated with PIN and PSOS, while PIN and PSOS are positively contemporaneously correlated. Although DY reported positive pairwise correlations between PIN, APIN and PSOS in the cross-section context across different stocks, our results show a different contemporaneous pattern for each stock in the time-series context. On a daily basis, a strong information signal increases APIN and reduces common shocks caused by disagreement in information interpretation, hence causing negative contemporaneous correlation between APIN and PSOS. On the other hand, PSOS is a component of the PIN measure, as argued by DY, hence inducing negative contemporaneous correlation between APIN and PIN.

5. Conclusions

In this paper we propose a method to estimate time varying APIN and PSOS suggested by DY as measures of asymmetric information and illiquidity, respectively. Our method is an extension of TTTW using high-frequency transaction data, which is based on an AACD model of expected durations of buy- and sell-orders. We allow the expected duration of buy- and sell-orders to be dependent on covariates such as lagged duration, lagged conditional expected duration, lagged trade direction and lagged trade volume. Also, we incorporate into our model time varying probabilities of no news, good news, bad news and symmetric order-flow shock. The model parameters are estimated using MLE, from which we obtain daily estimates of APIN and PSOS. The results provide an enhanced methodology to study the effects of asymmetric information and illiquidity on asset pricing. Our empirical results indicate that daily APIN is more stable than daily PIN. PSOS is correlated with average daily effective spread and daily volatility, supporting that it is a measure of illiquidity. We also observe the interesting result that the daily PSOS series exhibit a sporadic pattern of extended periods of no common shocks intermingled with clustered periods of active common-shock trading.

References

Aktas N, de Bodt E, Declerck F, Oppens HV. 2007. The *PIN* anomaly around M&A announcements. *Journal of Financial Markets* **10**: 169–191.

Amihud A. 2002. Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets* **5**: 31–56.

Andersen TG. 1996. Return volatility and trading volume: An information flow interpretation of stochastic volatility. *Journal of Finance* **51**: 169–204.

Bauwens L, Giot P. 2000. The logarithmic ACD model: An application to the bid-ask quote process of three NYSE stocks. *Annales d'Économie et de Statistique* **60**: 117–149.

Bauwens L, Giot P. 2003. Asymmetric ACD models: Introducing price information in ACD models. *Empirical Economics* **28**: 709–731.

Bauwens L, Giot P, Grammig J, Veredas D. 2004. A comparison of financial duration models via density forecast. *International Journal of Forecasting* **20**: 589–604.

Berry TD, Howe KM. 1994. Public information arrival. *Journal of Finance* **49**: 1331–1346.

Diamond DW, Verrecchia R. 1991. Disclosure, liquidity, and the cost of capital. *Journal of Finance* **66**: 1325–1359.

Drost FC, Werker BJM. 2004. Semiparametric duration models. *Journal of Business and Economic Statistics* **22**: 40–50.

Duarte J, Young L. 2009. Why is PIN priced? *Journal of Financial Economics* **91**: 119–138.

Easley D, Hvidkjaer S, O'Hara M. 2002. Is information risk a determinant of asset returns? *Journal of Finance* **57**: 2185–2221.

Easley D, Kiefer NM, O'Hara M. 1996. Cream-skimming or profit-sharing? The curious role of purchased order flow. *Journal of Finance* **51**: 811–833.

Easley D, Kiefer NM, O'Hara M. 1997. The information content of the trading process. *Journal of Empirical Finance* **4**: 159–186.

Easley D, O'Hara M. 2004. Information and the cost of capital. *Journal of Finance* **59**: 1553–1583.

Engle RF, Russell JR. 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* **66**: 1127–1162.

Hui EC, Wang Z, Yiu CK, Wong H. 2011. Informed trading in real estate stock market. Working paper. Polytechnic University of Hong Kong.

Lee CMC, Ready MJ. 1991. Inferring trade direction from intraday data. *Journal of Finance* **46**: 733–746.

Li J, Wu C. 2006. Daily return volatility, bid-ask spreads, and information flow: Analyzing the information content of volume, *Journal of Business* **79**: 2697–2739.

Mitchell ML, Mulherin JH. 1994. The impact of public information on the stock market. *Journal of Finance* **49**: 923–950.

Pacurar M. 2008. Autoregressive conditional duration models in finance: A survey of the theoretical and empirical literature. *Journal of Economic Surveys* **22**: 711–751.

Sarkar A, Schwarts RA. 2009. Market sideness: Insights into motives trade initiation. *Journal of Finance* **64**: 375–423.

Tay, A, Ting C, Tse YK, Warachka M. 2009. Using high-frequency transaction data to estimate the probability of informed trading. *Journal of Financial Econometrics* **7**: 288–311.

Tay, A, Ting C, Tse YK, Warachka M. 2011. The impact of transaction duration, volume, and direction on price dynamics and volatility. *Quantitative Finance* **11**: 447–457.

Tse YK, Yang TT. 2012. Estimation of high-frequency volatility: An autoregressive conditional duration approach. *Journal of Business and Economic Statistics*, forthcoming.

Table 1. Conditional expected durations for the APIN-AACD model

Buy-initiated trade	Sell-initiated trade
$\psi_{1i}^{CB} = \psi_{1i}^{N} e^{-\mu_{CB}}$	$\psi_{-1,i}^{CB} = \psi_{-1,i}^{N} e^{-(\mu_B+\mu_{CB})}$
$\psi_{1i}^{B} = \psi_{1i}^{N}$	$\psi_{-1,i}^{B} = \psi_{-1,i}^{N} e^{-\mu_B}$
$\psi_{1i}^{CG} = \psi_{1i}^{N} e^{-(\mu_G+\mu_{CG})}$	$\psi_{-1,i}^{CG} = \psi_{-1,i}^{N} e^{-\mu_{CG}}$
$\psi_{1i}^{G} = \psi_{1i}^{N} e^{-\mu_G}$	$\psi_{-1,i}^{G} = \psi_{-1,i}^{N}$
$\psi_{1i}^{CN} = \psi_{1i}^{N} e^{-\mu_{CN}}$	$\psi_{-1,i}^{CN} = \psi_{-1,i}^{N} e^{-\mu_{CN}}$

Notes: ψ_{ji}^{N} is defined in Equation (3.2). The positive constants μ_{CB}, μ_B, μ_{CG}, μ_G and μ_{CN} are unknown parameters.

Table 2. Summary statistics of the data, the EHO-PIN results and the DY-APIN-PSOS results

	IBM	GE	PG	WMT
Panel A: Summary statistics				
Frequency of buy-orders (%)	47.81	45.38	47.85	44.98
Frequency of sell-orders (%)	52.19	54.62	52.15	55.03
Serial correlation of trade direction	0.11	0.16	0.13	0.10
Runs test of trade direction	0.00	0.00	0.00	0.00
LB(15) for raw durations ($\times 10^5$)	5.69	7.67	4.95	6.06
LB(15) for DA durations ($\times 10^5$)	3.55	5.67	3.09	4.07
Average logarithmic trade volume	5.87	6.56	6.08	6.29
Average trade volume	712.71	2,149.12	1,038.07	1,366.44
Average daily number of trades	4,806.28	5,419.49	4,468.32	5,182.37
Average daily number of buy-orders	2,297.68	2,459.20	2,138.23	2,330.77
Average daily number of sell-orders	2,508.60	2,960.29	2,330.09	2,851.60
Number of observations	3,623,936	4,086,293	3,369,114	3,907,509
Panel B: Estimated EHO-PIN results				
π_G	0.077	0.116	0.035	0.165
π_N	0.750	0.717	0.685	0.727
π_B	0.174	0.167	0.280	0.109
EHO-PIN	0.069	0.083	0.081	0.061
Panel C: Estimated DY-APIN-PSOS results				
π_G	0.147	0.106	0.045	0.106
π_{CG}	0.045	0.037	0.023	0.041
π_N	0.438	0.474	0.333	0.414
π_{CN}	0.134	0.167	0.167	0.159
π_B	0.180	0.160	0.287	0.202
π_{CB}	0.055	0.056	0.144	0.078
APIN	0.057	0.050	0.071	0.052
PSOS	0.124	0.146	0.154	0.121

Notes: “Serial correlation of trade direction” is the sample autocorrelation at lag 1. “Runs test of trade direction” is the p-value of the Wald-Wolfowitz test for randomness for trade direction. LB(15) is the Ljung-Box statistic with 15 lags.

Table 3. Estimation results for the PIN- and APIN-AACD models

	IBM				GE			
	PIN-AACD		APIN-AACD		PIN-AACD		APIN-AACD	
γ_1	-0.773	(0.085)	-0.262	(0.053)	-0.853	(0.091)	-0.911	(0.073)
γ_2	3.286	(0.391)	1.524	(0.126)	2.902	(0.353)	1.210	(0.210)
γ_3	3.538	(0.406)	10.819	(1.000)	2.145	(0.361)	9.669	(1.242)
γ_4	7.245	(0.703)	11.795	(0.927)	6.230	(0.722)	10.367	(1.217)
γ_5			5.285	(0.431)			6.319	(0.666)
$\nu_{1,-1}$	0.125	(0.002)	0.122	(0.002)	0.263	(0.002)	0.267	(0.002)
$\nu_{1,1}$	0.099	(0.002)	0.093	(0.002)	0.221	(0.003)	0.222	(0.003)
β_1	0.899	(0.001)	0.903	(0.001)	0.777	(0.002)	0.774	(0.001)
α_1	0.073	(0.001)	0.072	(0.001)	0.122	(0.001)	0.122	(0.001)
ς_1	-0.005	(0.000)	-0.005	(0.000)	-0.014	(0.000)	-0.014	(0.000)
$\nu_{-1,-1}$	0.041	(0.001)	0.110	(0.002)	0.093	(0.001)	0.217	(0.003)
$\nu_{-1,1}$	0.055	(0.001)	0.107	(0.002)	0.102	(0.001)	0.181	(0.002)
β_{-1}	0.957	(0.001)	0.895	(0.001)	0.904	(0.001)	0.793	(0.002)
α_{-1}	0.042	(0.000)	0.067	(0.001)	0.084	(0.001)	0.114	(0.001)
ς_{-1}	0.002	(0.000)	0.006	(0.000)	0.006	(0.000)	0.015	(0.000)
μ_B	0.084	(0.003)	0.180	(0.002)	0.123	(0.004)	0.236	(0.003)
μ_G	0.216	(0.002)	0.095	(0.002)	0.319	(0.002)	0.173	(0.002)
μ_{CB}			0.233	(0.003)			0.361	(0.002)
μ_{CG}			0.279	(0.003)			0.271	(0.004)
μ_{CN}			0.199	(0.002)			0.277	(0.002)
LB(15) ($\times 10^3$)	8.058		6.330		7.510		20.594	
AIC ($\times 10^7$)	1.21681		1.21558		1.34778		1.34520	
SBC ($\times 10^7$)	1.21683		1.21561		1.34780		1.34522	

	PG				WMT			
	PIN-AACD		APIN-AACD		PIN-AACD		APIN-AACD	
γ_1	-0.724	(0.075)	-0.518	(0.073)	-0.921	(0.088)	-0.629	(0.030)
γ_2	1.303	(0.101)	0.017	(0.184)	1.480	(0.291)	0.550	(0.111)
γ_3	4.255	(0.337)	11.791	(2.018)	12.081	(5.042)	13.659	(0.774)
γ_4	6.008	(0.317)	13.165	(2.067)	13.570	(8.427)	13.997	(0.647)
γ_5			3.757	(0.378)			3.8834	(0.422)
$\nu_{1,-1}$	0.191	(0.003)	0.183	(0.003)	0.208	(0.010)	0.223	(0.004)
$\nu_{1,1}$	0.146	(0.003)	0.137	(0.004)	0.082	(0.012)	0.095	(0.003)
β_1	0.843	(0.002)	0.851	(0.003)	0.868	(0.006)	0.856	(0.004)
α_1	0.092	(0.001)	0.090	(0.001)	0.091	(0.007)	0.094	(0.004)
ς_1	-0.008	(0.000)	-0.007	(0.000)	-0.000	(0.008)	-0.000	(0.002)
$\nu_{-1,-1}$	0.043	(0.001)	0.110	(0.004)	0.058	(0.009)	0.075	(0.002)
$\nu_{-1,1}$	0.080	(0.001)	0.140	(0.009)	0.086	(0.021)	0.102	(0.003)
β_{-1}	0.945	(0.001)	0.881	(0.007)	0.928	(0.012)	0.911	(0.002)
α_{-1}	0.051	(0.001)	0.074	(0.002)	0.063	(0.012)	0.069	(0.002)
ς_{-1}	0.002	(0.000)	0.005	(0.000)	0.002	(0.057)	0.003	(0.002)
μ_B	0.089	(0.003)	0.199	(0.002)	0.088	(0.004)	0.094	(0.003)
μ_G	0.250	(0.002)	0.111	(0.003)	0.223	(0.004)	0.180	(0.002)
μ_{CB}			0.302	(0.006)			0.158	(0.004)
μ_{CG}			0.346	(0.010)			0.184	(0.002)
μ_{CN}			0.204	(0.004)			0.111	(0.003)
LB(15) ($\times 10^3$)	5.285		6.618		8.047		7.947	
AIC ($\times 10^7$)	1.12973		1.12849		1.30381		1.30307	
SBC ($\times 10^7$)	1.12976		1.12851		1.30383		1.30310	

Notes: AIC is the Akaike Information Criterion. SBC is Schwarz Bayesian Criterion. Figures within parentheses are standard errors. LB(15) is the Ljung-Box statistic with 15 lags.

Table 4. Correlations between daily spread/variance/return and PIN/APIN/PSOS

	IBM			GE		
	PIN	APIN	PSOS	PIN	APIN	PSOS
Effective spread	0.046	0.011	0.107	0.122	0.049	0.101
	(0.205)	(0.765)	(0.003)	(0.001)	(0.179)	(0.006)
Variance	0.379	-0.129	0.382	0.493	-0.153	0.490
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Return	-0.101	-0.036	-0.128	0.055	-0.152	0.011
	(0.006)	(0.319)	(0.000)	(0.129)	(0.000)	(0.754)
PSOS	0.884	-0.623		0.878	-0.657	
	(0.000)	(0.000)		(0.000)	(0.000)	
APIN	-0.472			-0.405		
	(0.000)			(0.000)		
	$\hat{\pi}_G$	$\hat{\pi}_B$		$\hat{\pi}_G$	$\hat{\pi}_B$	
Return	0.148	-0.109		0.303	-0.285	
	(0.000)	(0.003)		(0.000)	(0.000)	

	PG			WMT		
	PIN	APIN	PSOS	PIN	APIN	PSOS
Effective spread	0.018	0.006	0.068	0.144	–0.010	0.091
	(0.623)	(0.869)	(0.062)	(0.000)	(0.790)	(0.013)
Variance	0.142	-0.126	0.231	0.231	0.030	0.135
	(0.000)	(0.001)	(0.000)	(0.000)	(0.417)	(0.000)
Return	0.124	-0.115	-0.021	0.117	0.064	0.040
	(0.001)	(0.002)	(0.567)	(0.001)	(0.078)	(0.274)
PSOS	0.740	-0.767		0.799	-0.745	
	(0.000)	(0.000)		(0.000)	(0.000)	
APIN	-0.874			-0.228		
	(0.000)			(0.000)		
	$\hat{\pi}_G$	$\hat{\pi}_B$		$\hat{\pi}_G$	$\hat{\pi}_B$	
Return	0.208	-0.182		0.168	-0.195	
	(0.000)	(0.000)		(0.000)	(0.000)	

Notes: Figures within parentheses are p-values (null hypothesis is zero-correlation). $\hat{\pi}_G$ and $\hat{\pi}_B$ are the estimated daily probabilities of good news and bad news, respectively, for the APIN-AACD model.

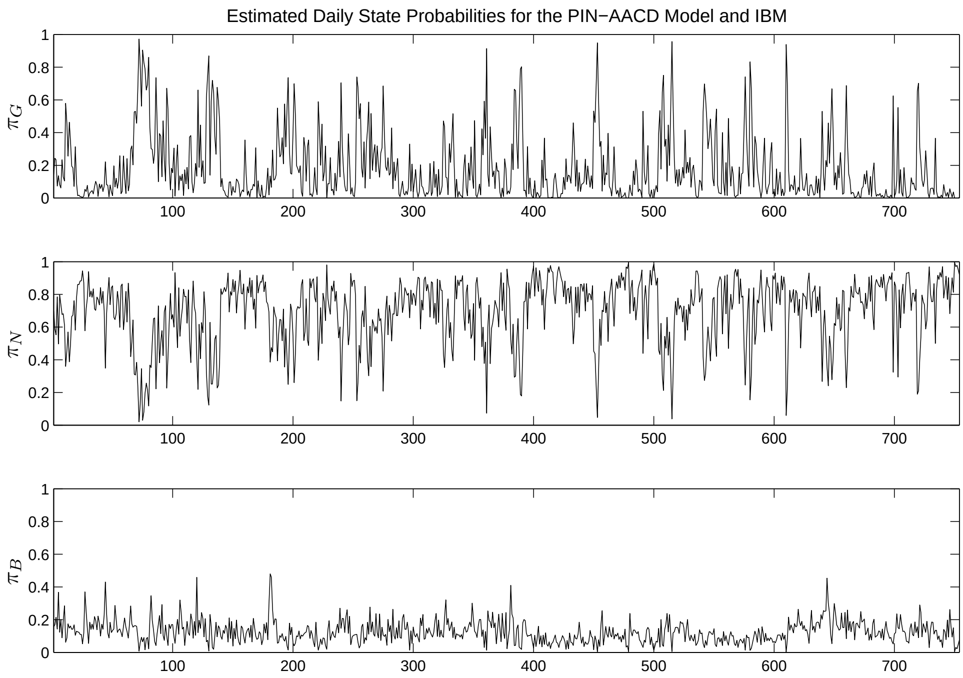

FIGURE 1. Model-implied probabilities for the PIN-AACD model of IBM.

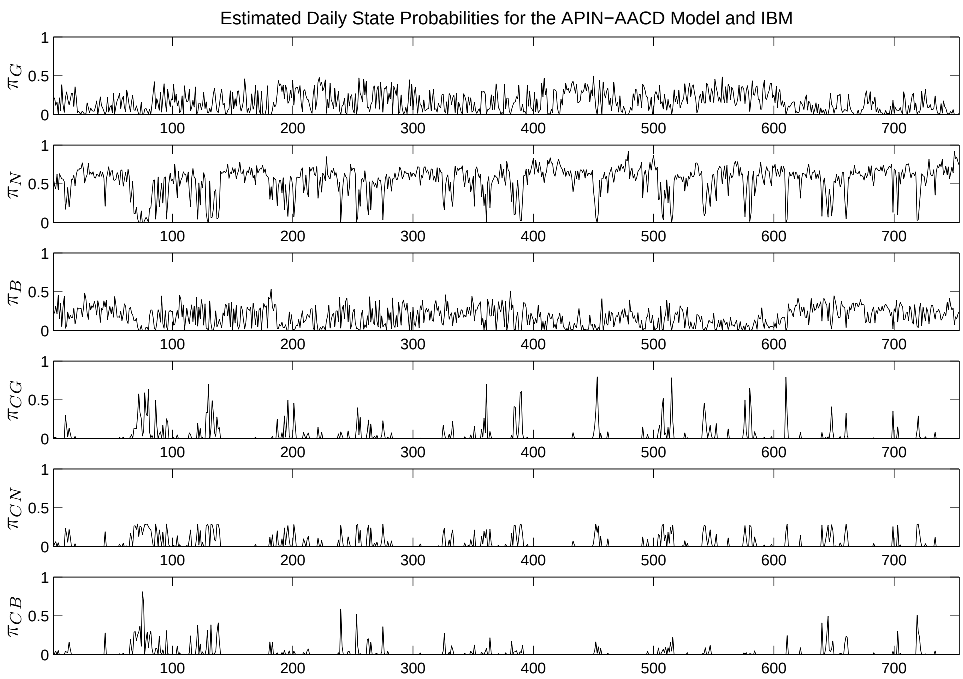

FIGURE 2. Model-implied probabilities for the APIN-AACD model of IBM.

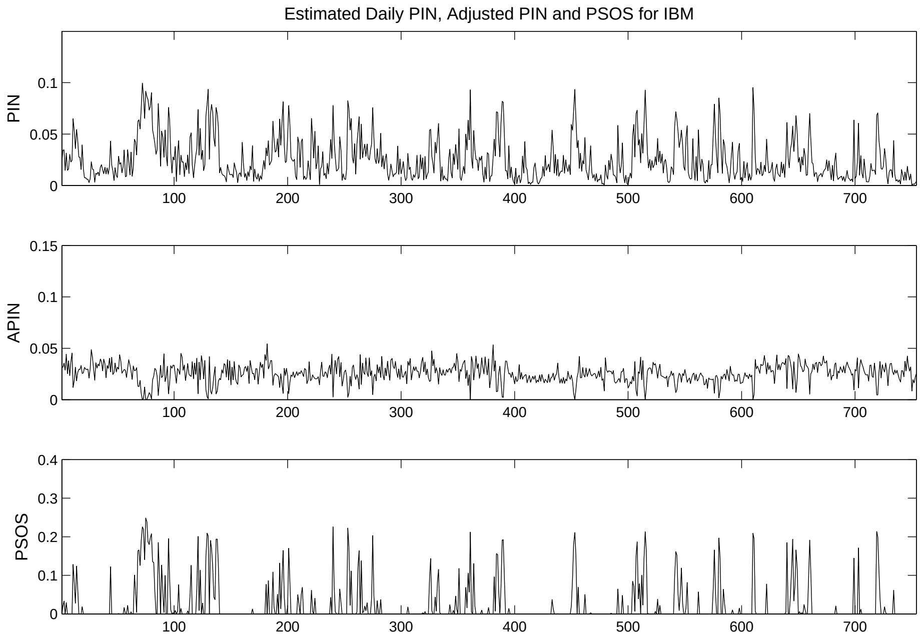

Figure 3. Estimated Daily PIN, APIN and PSOS for IBM.

Estimation of Time Varying Adjusted Probability of Informed Trading and Probability of Symmetric Order-Flow Shock

Web Appendix

Daniel Preve, City University of Hong Kong
Yiu-Kuen Tse, Singapore Management University

May 2012

Easley *et al.* (2002, EHO) proposed a market microstructure model to derive a measure of asymmetric information reflecting the relative intensity of informed versus uninformed (liquidity) trades, called the probability of informed trading, PIN. As described in Figure 1, the PIN model assumes that each trading day may be classified as one with news or no news. Furthermore, a day with news can be one with good news or bad news. The daily aggregate number of buyer- and seller-initiated trades (buy and sell orders) are assumed to follow independent Poisson distributions with intensities dependent on whether the trading day is one with good news, bad news or no news. In the model there are two types of traders, informed traders who trade based on relevant news or information, and uninformed traders who trade for reasons not accounted for by relevant information, such as portfolio rebalancing and liquidity needs.

Let B_d and S_d denote the aggregate number of buy- and sell-orders on day d, respectively. In the PIN model, B_d and S_d are assumed to be independent Poisson random variables, with different intensities for days with bad news (B), good news (G) and no news (N). Let θ_E denote the probability of news being released on day d and let θ_B denote the probability of bad news, conditional on the release of news. Thus, the daily state probabilities are $\pi_B = \theta_E\theta_B$, $\pi_G = \theta_E(1 - \theta_B)$ and $\pi_N = 1 - \theta_E$, for a day with bad news, good news and no news, respectively. The means of B_d and S_d (the intensity parameters) vary according to whether the trading day is one with good news, bad news or no news. In particular, for a day with no news, the means of B_d and S_d are λ_1 and λ_{-1}, respectively. For a day with bad news the sell intensity increases by a constant δ, while the buy intensity remains the same as for a day with no news. Similarly, for a day with good news the buy intensity increases by δ, while the sell intensity stays the same as for a no-news day. The PIN model assumes that orders due to informed and uninformed traders are independent.

For each trading day d, the joint distribution of B_d and S_d is given by

$$f(B_d, S_d) = \sum_{s \in \mathcal{S}} f(B_d, S_d, s) = \sum_{s \in \mathcal{S}} \pi_s f(B_d, S_d | s), \qquad \text{for } \mathcal{S} = \{B, G, N\},$$

implying that the daily expected total number of trades is

$$\begin{aligned}
\mathrm{E}(B_d + S_d) &= \sum_{s \in \mathcal{S}} \pi_s \mathrm{E}(B_d + S_d | s) \\
&= \theta_E \theta_B (\lambda_1 + \lambda_{-1} + \delta) + \theta_E (1 - \theta_B)(\lambda_1 + \lambda_{-1} + \delta) + (1 - \theta_E)(\lambda_1 + \lambda_{-1}) \\
&= \underbrace{\lambda_1 + \lambda_{-1}}_{\mathrm{P}_1} + \underbrace{\theta_E \delta}_{\mathrm{P}_2},
\end{aligned}$$

where P_1 and P_2 are the expected numbers of trades due to uninformed and informed traders, respectively. EHO compute the PIN as the relative intensity of informed trades to the intensity of all trades, so that

$$\mathrm{PIN} = \frac{\mathrm{P}_2}{\mathrm{P}_1 + \mathrm{P}_2}.$$

The parameters in the PIN model can be estimated using the MLE method. With D days of data, the likelihood function is

$$\prod_{d=1}^{D} \left[\pi_B \frac{\lambda_1^{B_d}}{B_d!} e^{-\lambda_1} \frac{(\lambda_{-1} + \delta)^{S_d}}{S_d!} e^{-(\lambda_{-1} + \delta)} + \pi_G \frac{(\lambda_1 + \delta)^{B_d}}{B_d!} e^{-(\lambda_1 + \delta)} \frac{\lambda_{-1}^{S_d}}{S_d!} e^{-\lambda_{-1}} + \pi_N \frac{\lambda_1^{B_d}}{B_d!} e^{-\lambda_1} \frac{\lambda_{-1}^{S_d}}{S_d!} e^{-\lambda_{-1}} \right],$$

as given by EHO.

Duarte and Young (2009, DY) extended the PIN model of EHO by allowing for the arrival rate of informed sellers to be different from the arrival rate of informed buyers and, more importantly, by allowing both buy- and sell-order flows to increase on certain days even when there is no news. Their APIN model, outlined in Figure 2, has three additional states representing days in which both the numbers of buys and sells increase due to symmetric order-flow shocks, or common shocks for short. The motivation for the first extension is to improve the ability of the PIN model to account for the fact that buy-order flow has a larger variance than sell-order flow, for almost all firms, in their empirical study. The second, more important, extension allows for increased buy and sell variations, and a positive correlation between buys and sells, as each day a common shock may occur that causes both buy- and sell-order flows to increase.

In the APIN model B_d and S_d have different intensities for days with bad news and a common shock (CB), good news and a common shock (CG) and no news and a common shock (CN). The occurrence of a common shock is assumed to be independent of the arrival of news (good, bad or no news). Let θ_C denote the daily probability of a common shock. The state space $\mathcal{S}$ then represents cases of no common shocks, and the extended state space is $\mathcal{S}^* = \{CB, CG, CN, B, G, N\}$. In the event of a common shock, the buy intensity increases by $\triangle_1$ and the sell intensity by $\triangle_{-1}$. Possible causes for common shocks include the arrival of public news the implications of which traders disagree, and coordinated trading on certain days in order to reduce trading costs (Duarte and Young, 2009). The APIN model also allows the arrival rate of informed sellers to be different from the arrival rate of informed buyers. On a day with bad news the sell intensity increases by δ_{-1}, while the buy intensity remains the same as for a day with no news. On a day with good news the buy intensity increases by δ_1, while the sell intensity stays the same as for a no-news day.

The likelihood function for the APIN model is given by

$$\prod_{d=1}^{D}\Bigg[\pi_{CB}\frac{(\lambda_1+\triangle_1)^{B_d}}{B_d!}e^{-(\lambda_1+\triangle_1)}\frac{(\lambda_{-1}+\delta_{-1}+\triangle_{-1})^{S_d}}{S_d!}e^{-(\lambda_{-1}+\delta_{-1}+\triangle_{-1})}$$
$$+\pi_B\frac{\lambda_1^{B_d}}{B_d!}e^{-\lambda_1}\frac{(\lambda_{-1}+\delta_{-1})^{S_d}}{S_d!}e^{-(\lambda_{-1}+\delta_{-1})}$$
$$+\pi_{CG}\frac{(\lambda_1+\delta_1+\triangle_1)^{B_d}}{B_d!}e^{-(\lambda_1+\delta_1+\triangle_1)}\frac{(\lambda_{-1}+\triangle_{-1})^{S_d}}{S_d!}e^{-(\lambda_{-1}+\triangle_{-1})}+\pi_G\frac{(\lambda_1+\delta_1)^{B_d}}{B_d!}e^{-(\lambda_1+\delta_1)}\frac{\lambda_{-1}^{S_d}}{S_d!}e^{-\lambda_{-1}}$$
$$+\pi_{CN}\frac{(\lambda_1+\triangle_1)^{B_d}}{B_d!}e^{-(\lambda_1+\triangle_1)}\frac{(\lambda_{-1}+\triangle_{-1})^{S_d}}{S_d!}e^{-(\lambda_{-1}+\triangle_{-1})}+\pi_N\frac{\lambda_1^{B_d}}{B_d!}e^{-\lambda_1}\frac{\lambda_{-1}^{S_d}}{S_d!}e^{-\lambda_{-1}}\Bigg].$$

For this model it is straightforward to show that the expected value of all trades for day d can be decomposed into three parts

$$\mathrm{E}(B_d+S_d)=\underbrace{\lambda_1+\lambda_{-1}}_{\mathrm{P}_1}+\underbrace{\theta_E[(1-\theta_B)\delta_1+\theta_B\delta_{-1}]}_{\mathrm{P}_2}+\underbrace{\theta_C(\triangle_1+\triangle_{-1})}_{\mathrm{P}_3}.$$

Tay *et al.* (2009, TTTW) estimated the PIN model assuming the latent trade directions follow the AACD model, introducing the PIN-AACD model. The conditional intensities of buy- and sell-orders under different news environments are illustrated in Figure 3. In a similar manner, Figure 4 illustrates the extended APIN-AACD model, in which the DY APIN model is estimated using the AACD model.

Figures 5, 6 and 7 present additional graphical plots of the estimated daily probabilities of news as well as PIN/APIN/PSOS for the GE stock. Tables 1 and 2 present some summary statistics of the PIN-AACD and APIN-AACD models for the four stocks, respectively. The Matlab codes for the computation of the PIN-AACD and APIN-AACD models can be downloaded, the user guide of which is included in this appendix.

TABLE 1. Summary statistics of the PIN-AACD model with time varying probabilities.

	IBM	GE	PG	WMT
π_G				
Mean	0.182	0.173	0.156	0.139
Std Dev	0.203	0.184	0.116	0.124
π_N				
Mean	0.692	0.704	0.685	0.725
Std Dev	0.201	0.175	0.099	0.104
π_B				
Mean	0.126	0.123	0.155	0.136
Std Dev	0.069	0.065	0.068	0.090
PIN				
Mean	0.025	0.035	0.028	0.022
Std Dev	0.021	0.025	0.013	0.011

TABLE 2. Summary statistics of the APIN-AACD model with time varying probabilities.

	IBM	GE	PG	WMT
π_G				
Mean	0.165	0.114	0.152	0.140
Std Dev	0.127	0.085	0.120	0.102
π_N				
Mean	0.543	0.653	0.565	0.601
Std Dev	0.191	0.206	0.132	0.162
π_B				
Mean	0.182	0.118	0.185	0.158
Std Dev	0.122	0.087	0.128	0.105
π_{CG}				
Mean	0.042	0.028	0.022	0.024
Std Dev	0.115	0.076	0.056	0.066
π_{CN}				
Mean	0.041	0.067	0.062	0.059
Std Dev	0.083	0.136	0.131	0.125
π_{CB}				
Mean	0.027	0.020	0.015	0.018
Std Dev	0.082	0.063	0.041	0.055
APIN				
Mean	0.026	0.026	0.029	0.021
Std Dev	0.009	0.009	0.010	0.006
PSOS				
Mean	0.026	0.033	0.025	0.014
Std Dev	0.056	0.068	0.052	0.031

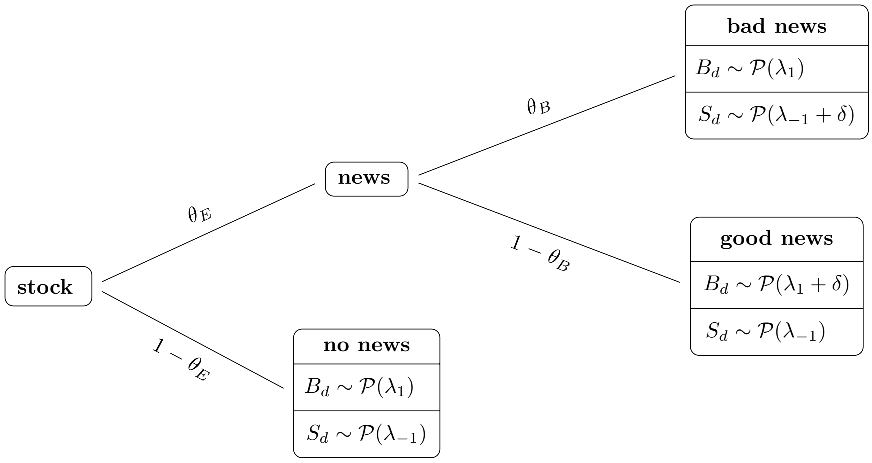

FIGURE 1. Trading tree for the PIN model: B_d and S_d are the total number of buy and sell orders on trading day d, respectively. We write $B_d \sim \mathcal{P}(\lambda_1)$ to indicate that B_d is Poisson distributed with intensity parameter (mean and variance) λ_1. On each trading day news arrive with probability θ_E. On a no-news day, B_d is Poisson distributed with intensity λ_1 and S_d is Poisson distributed with intensity λ_{-1}. Bad news causes an increase in the intensity of S_d, consequently S_d is Poisson distributed with intensity $\lambda_{-1} + \delta$ on a bad-news day. Similarly, B_d is Poisson distributed with intensity $\lambda_1 + \delta$ on a good-news day.

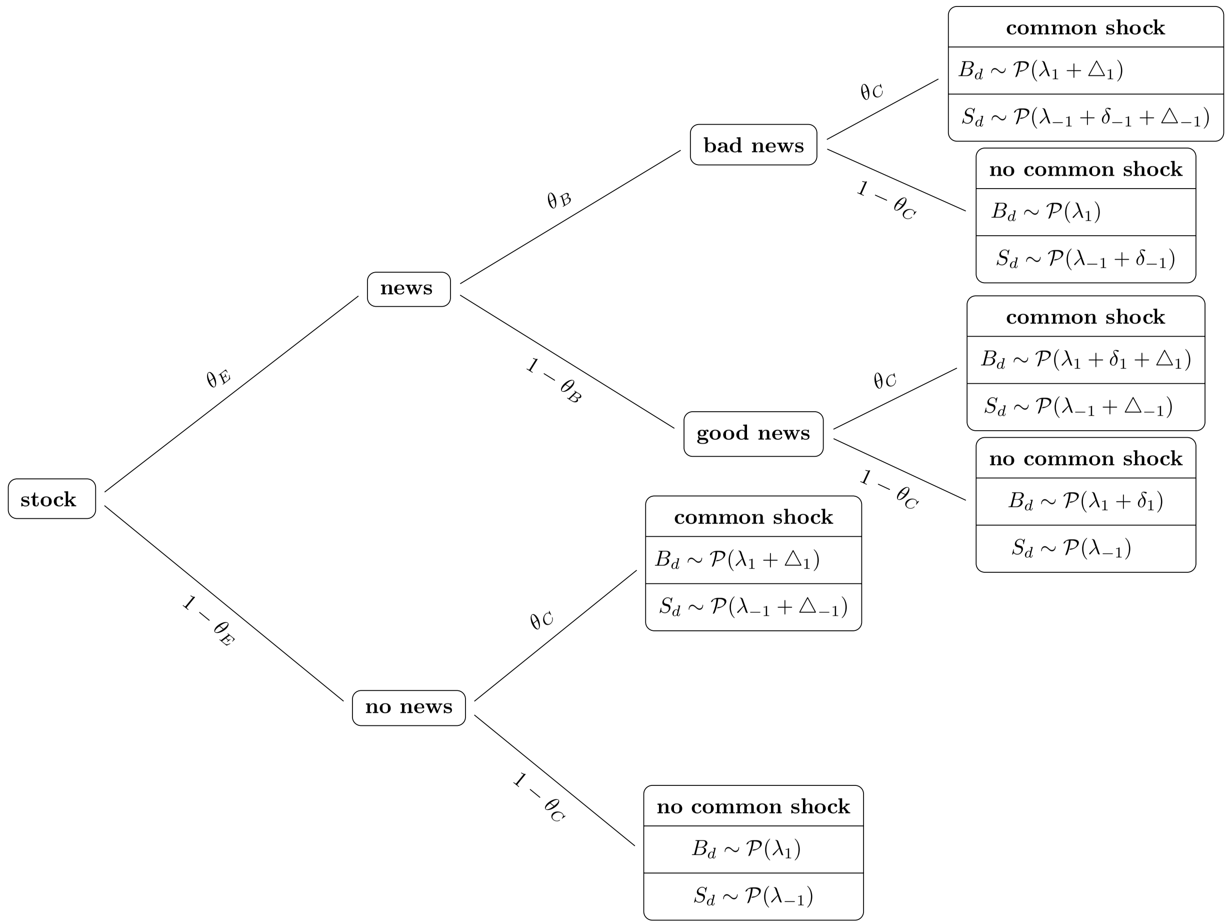

FIGURE 2. Trading tree for the APIN model: B_d and S_d are the total number of buy and sell orders on trading day d, respectively. We write $B_d \sim \mathcal{P}(\lambda_1)$ to indicate that B_d is Poisson distributed with intensity parameter (mean and variance) λ_1. In contrast to the PIN model, the APIN model has three additional states representing trading days in which both the intensity of B_d and S_d increase due to common shocks. These days occur with probability θ_C, irrespective of no news, good news or bad news.

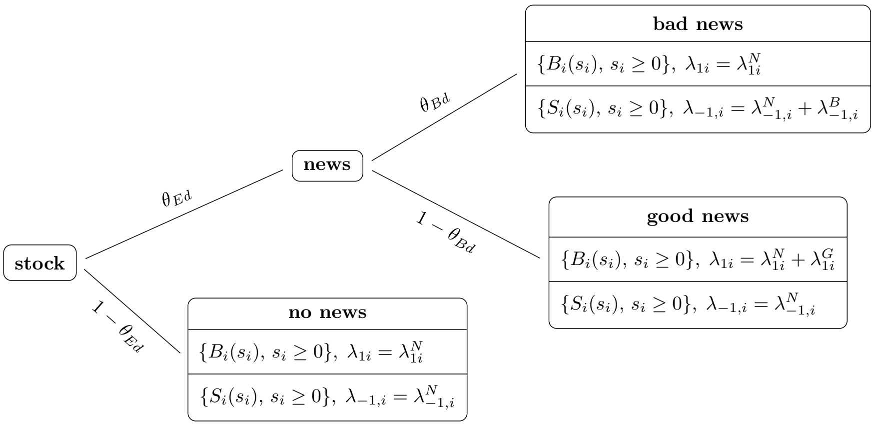

FIGURE 3. Trading tree for the PIN-AACD model: $\{B_i(s_i),\, s_i \geq 0\}$ and $\{S_i(s_i),\, s_i \geq 0\}$ are the latent Poisson processes of buy and sell orders initiated at time t_{i-1} on trading day d, respectively, given the information Φ_{i-1}. On each trading day news arrive with probability θ_{Ed}. On a no-news day the conditional intensity of the buy orders is λ_{1i}^N and the conditional intensity of the sell orders is $\lambda_{-1,i}^N$. On a bad-news day the conditional intensity of sell orders increase by $\lambda_{-1,i}^B$, while that of buy orders remains the same as on a no-news day. Similarly, on a good-news day the conditional intensity of buy orders increase by λ_{1i}^G, while that of sell orders remains the same as on a no-news day.

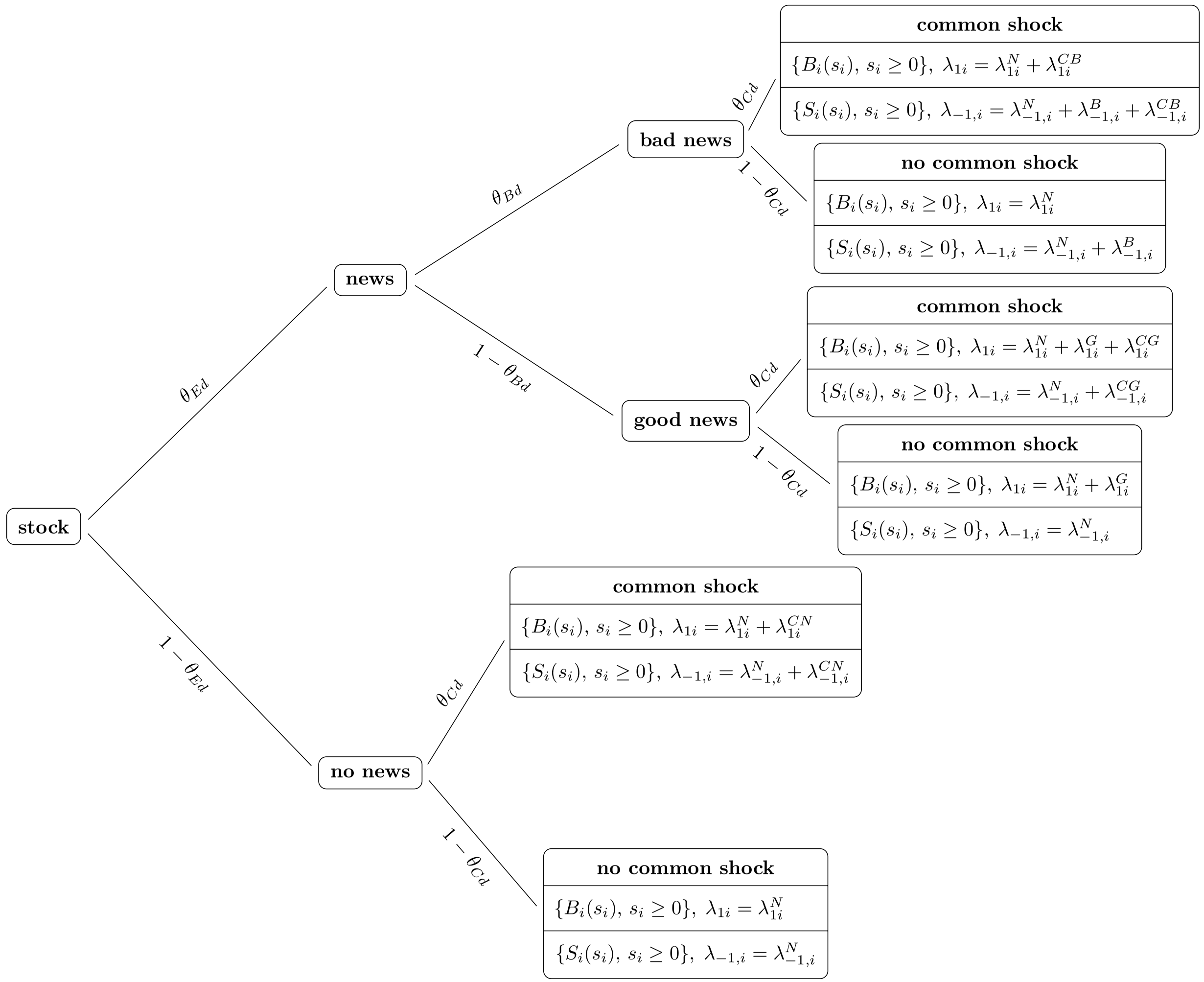

FIGURE 4. Trading tree for the APIN-AACD model: $\{B_i(s_i),\ s_i \geq 0\}$ and $\{S_i(s_i),\ s_i \geq 0\}$ are the latent Poisson processes of buy and sell orders initiated at time t_{i-1} on trading day d, respectively, given the information Φ_{i-1}. In contrast to the PIN-AACD model, the APIN-AACD model has three additional states representing trading days in which both the conditional intensity of buy and sell orders increase due to common shocks. These days occur with probabilities θ_{Cd}, irrespective of no news, good news or bad news.

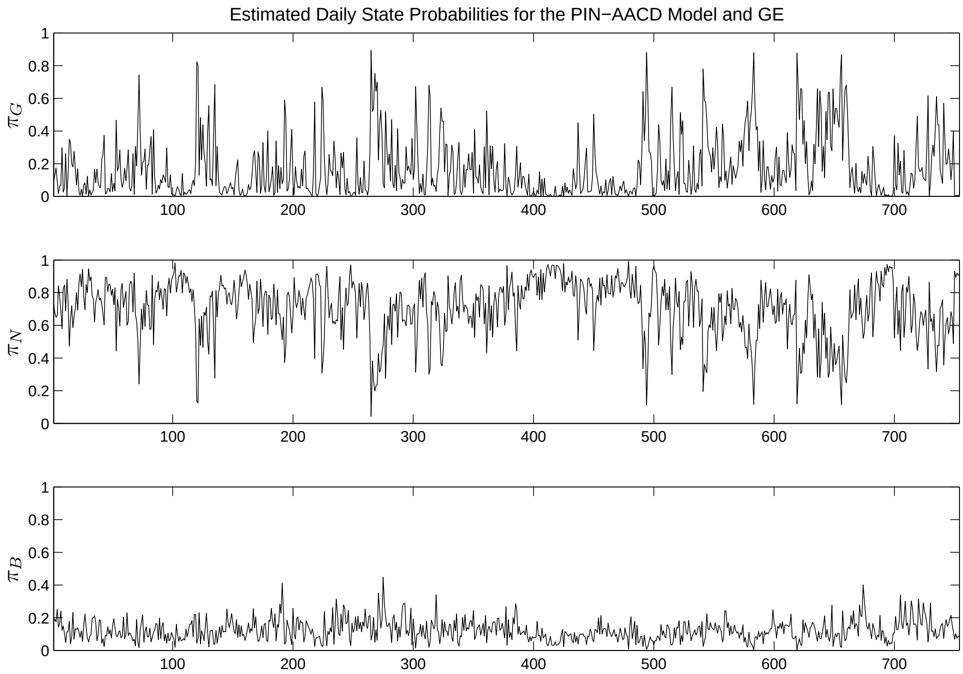

FIGURE 5. Model-implied probabilities for the PIN-AACD model of GE.

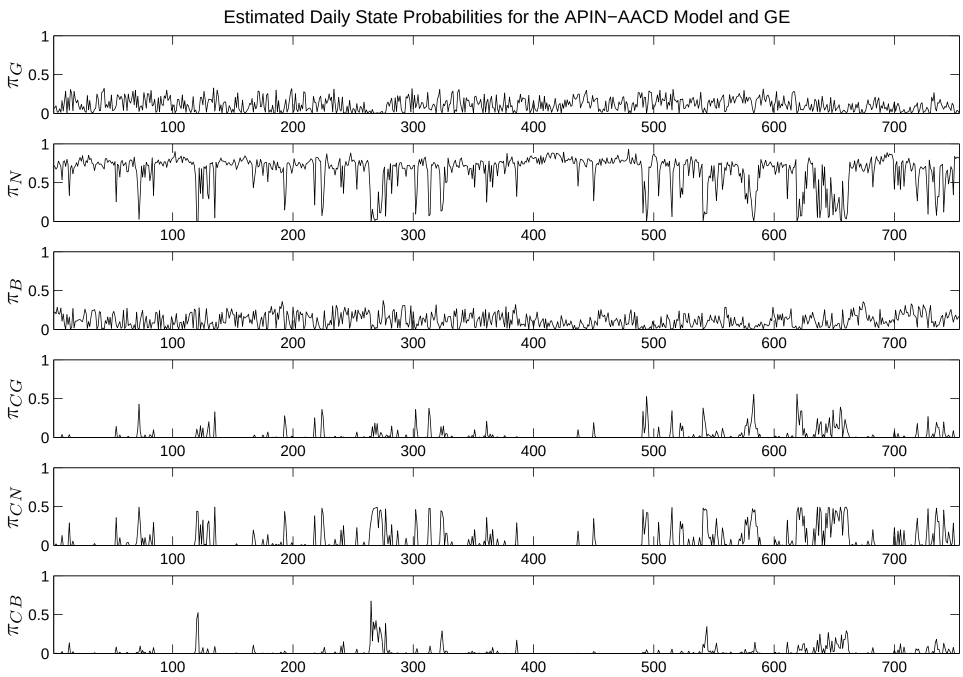

FIGURE 6. Model-implied probabilities for the APIN-AACD model of GE.

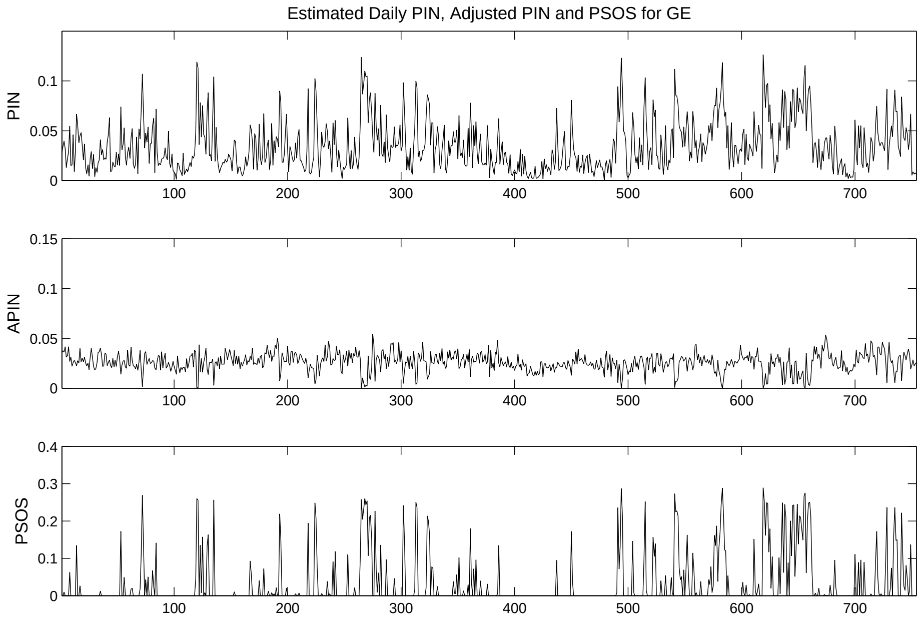

FIGURE 7. Estimated Daily PIN, APIN and PSOS for GE.

MATLAB CODES TO ACCOMPANY PREVE AND TSE (2012)

Introduction

This document summarizes the Matlab functions used in Preve and Tse (2012) for the estimation of the PIN- and APIN-AACD models and for the computation of the PIN, APIN and PSOS measures. These codes require Matlab 2009b (or later) and the optimization toolbox.

Overview

The list below shows the Matlab functions listed in alphabetical order. For more detailed information, type `help fun` in the command window to display a description of and syntax for the function `fun`.

- `computePIN0` – computes the daily estimated PIN for the PIN-AACD model
- `computePIN1` – computes the daily estimated APIN and PSOS for the APIN-AACD model
- `gaOptions0` – creates an options structure for the PIN-AACD model
- `gaOptions1` – creates an options structure for the APIN-AACD model
- `logL0` – log-likelihood function for the PIN-AACD model
- `logL1` – log-likelihood function for the APIN-AACD model
- `PIN0` – `computePIN0` support function
- `PIN1` – `computePIN1` support function

Daily PIN

The PIN-AACD model of Tay, Ting, Tse and Warachka (2009) can be estimated and its PIN measure computed in two steps:

(1) Run `gaOptions0`, then run `optimtool` with ga - Genetic Algorithm, `-logL0` as fitness function (number of variables=16, population size=250, generations=100, function tolerance=1e-8) and `fmincon` as hybrid function. Use 'Aineq', 'bineq' as linear inequalities, 'lb' and 'ub' as bounds and 'options' as options for `fmincon`. Export the result to a structure, e.g. IBM.Optimum.0.mat. Repeat step (1) ten times. Select the best optimum and use it in step (2).

(2) Using the optimum in step (1), run `computePIN0` to compute the daily estimated PIN for the PIN-AACD model with time-varying probabilities of news.

Daily APIN and PSOS

The APIN-AACD model of Preve and Tse (2012) can be estimated and its APIN and PSOS measures computed in two steps:

(1) Run `gaOptions1`, then run `optimtool` with ga - Genetic Algorithm, `-logL1` as fitness function (number of variables=20, population size=250, generations=100, function tolerance=1e-8) and `fmincon` as hybrid function. Use 'Aineq', 'bineq' as linear inequalities, 'lb' and 'ub' as bounds and 'options' as options for `fmincon`. Export the result to a structure, e.g. IBM.Optimum.1.mat. Repeat step (1) ten times. Select the best optimum and use it in step (2).

(2) Using the optimum in step (1), run `computePIN1` to compute the daily estimated adjusted PIN and PSOS for the APIN-AACD model with time-varying probabilities of news.