Finance Research Letters 31 (2019) 458–462

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

A note of techniques that mitigate floating-point errors in PIN estimation

Ke Wen-Chyan[a,*], Chen Hueiling[b], Lin Hsiou-Wei William[c]

[a] National Taipei University, No.151, University Rd., San-Shia District, New Taipei City 23741, Taiwan
[b] National Taiwan Normal University, No.162, Sec. 1, Heping E. Rd., Da'an Dist., Taipei City 10610, Taiwan
[c] National Taiwan University, No.1, Sec. 4, Roosevelt Rd., Taipei City 106, Taiwan

ARTICLE INFO

JEL classification:
C13
C60
G14

Keywords:
PIN
Maximum likelihood
Scaled trade counts
Floating-point exception

SUMMARY

This study aims at the estimation of the probability of informed trading (PIN), which may fail for stocks with high levels of trading activities due to a computer's floating-point exception (FPE). In this paper, we discuss two solutions of adopting scaled trade counts and reformulating the likelihood to estimate PIN for actively traded stocks. This study shows that, although scaled data mitigates the impact of the FPE, the effectiveness of scaled data, however, appears to under-perform when users adopt the unsuitable expression of the likelihood function. In contrast, the remedy of reformulating the likelihood is more stable.

1. Introduction

This study aims at a trade-off in adopting data-scaling versus likelihood-reformulating procedure for eliminating bias in deriving the probability of informed trading (PIN) measure, which is introduced in Easley and O'Hara (1987) and Easley et al. (1996), for stocks with high trading levels. Specifically, trading levels may be high in periods surrounding news announcements, such as firm earnings, seasoned public offerings, and mergers and acquisitions. Despite that the information asymmetry may be more pronounced during these periods, PIN estimation for frequently traded stocks, nevertheless, may fail due to a computer's floating-point exception (FPE; i.e., over/underflow).

Regardless of its wide acceptance and broad use, PIN measure attracts criticisms including FPE emerging during the estimation process. Invalidating PIN would substantially jeopardize the value added by the information asymmetry studies (e.g. Duarte and Young, 2009; Preve and Tse, 2012; Lai et al., 2014). Nevertheless, the debate on whether PIN is an appropriate measure of informed trading does not end. Strictly based on the unbiased PIN estimates, any statements over the effectiveness of PIN may be well grounded. Moreover, Akay et al. (2012) and Gan et al. (2017) argue that theoretically the assumptions of the PIN model fail to fit well the empirical data. Duarte et al. (2017) show that the PIN model is no more useful than an unsophisticated examination of market turnover in identifying private-information arrival. However, these studies demonstrate that PIN model may be imperfect but do not disprove the usefulness of estimation skills embedded in PIN, which, we argue, may be well adjusted to meet the needs of empiricists. When an improved PIN model is proposed, these skills may apply to the new setting. For example, Easley et al. (2012) develop VPIN (volume-synchronized PIN), which is estimated by the moment method. However, Ke and Lin (2017) model the VPIN metric and argue that Easley et al. (2012) VPIN metric is an information measure differing from PIN. Still applying the PIN estimation skills to

* Corresponding author.
E-mail addresses: wenchyan@gm.ntpu.edu.tw (W.-C. Ke), hlchen@ntnu.edu.tw (H. Chen), plin@ntu.edu.tw (H.-W.W. Lin).

https://doi.org/10.1016/j.frl.2018.12.017
Received 1 April 2018; Received in revised form 9 December 2018; Accepted 22 December 2018
Available online 25 December 2018

support their arguments, Ke and Lin (2017) adopt the maximum likelihood estimation (MLE) to derive the parameters. Namely, the skills may be applied to stabilize the computing.

Among the efforts exerted for enhancing the effectiveness of PIN, extant studies also propose remedial measures for an environment of high level of stock trading, which is one of the primary triggers of PIN estimate contamination and may accordingly distort PIN applications. Specifically, Lin and Ke (2011) show that large trade counts result in an underestimated PIN. Consistently, to avoid bias in estimating PIN, Aslan et al. (2011) adopt accounting and other market data to generate a proxy of PIN.

Lin and Ke (2011) propose a reformulated likelihood for mitigating FPE in PIN estimation. In contrast, Jackson (2013) provides a simple procedure that uses the scaled trade counts to estimate PIN for stocks with high trading levels but eliminates factoring the likelihood function. Namely, when the estimation fails for a given stock, researchers may equally scale the counts of buyer- and seller-initiated trades (*Buys* and *Sells*) and redo the estimation. The scaled observations help mitigate over/underflow problems, and the parameter estimates appear to be scaled back up.

However, based on Lin and Ke (2011), we analytically show that the remedy by using scaled trade counts may still be subject to FPE problem when adopting certain inappropriate expressions of likelihood function. Namely, although scaled data helps mitigate the impact of FPE, it fails to eliminate fully the FPE bias in PIN estimation because its effectiveness varies with the expression of the likelihood function.

The remainder is organized as follows. Section 2 shows that the scaled data do not effectively eliminate the FPE impact for an expression of the likelihood function prevalently adopted in prior studies. Section 3 specifies the suitable likelihood function for the data scaling approach of Jackson (2013), and Section 4 concludes this note.

2. Computing bias and scaled data

Drawing on Easley et al. (1996), *Buys* and *Sells* follow Poisson distributions. On day i, an information event occurs with the probability α. The given event is negative with the probability δ. Moreover, ε denotes the arrival rate of either uninformed *Buys* or *Sells*; and μ denotes the one of informed *Sells* (*Buys*) for a negative (positive) signal. Therefore, the joint probability density function (pdf) of (Bi, Si), with the elements being the observed numbers of *Buys* and *Sells*, respectively, is

$$f(B_i, S_i|\theta) \equiv \alpha\delta \mathrm{P}(B_i, \varepsilon)\mathrm{P}(S_i, \varepsilon + \mu) + \alpha(1-\delta)\mathrm{P}(B_i, \varepsilon + \mu)\mathrm{P}(S_i, \varepsilon) + (1-\alpha)\mathrm{P}(B_i, \varepsilon)\mathrm{P}(S_i, \varepsilon), \tag{1}$$

where $\mathrm{P}(x, \lambda) \equiv \exp(-\lambda)\lambda^x/x!$ is a Poisson pdf with mean λ evaluated at trade count x; and $\boldsymbol{\theta} \equiv (\alpha, \delta, \mu, \varepsilon)$ consists of these structural parameters.

With the assumption of independent daily information arrivals and given $\mathbf{T} \equiv \{(B_i, S_i)\}_{i=1}^{I}$, one may obtain the MLE estimate of $\boldsymbol{\theta}$ by solving the problem:

$$\max_{\theta\in\mathbf{BFS}} L_B(\theta|\mathbf{T}), \tag{2}$$

where $L_B(\boldsymbol{\theta}|\mathbf{T}) \equiv \sum_{i=1}^{I} L_B(\theta|B_i, S_i) = \sum_{i=1}^{I} \log[f(B_i, S_i|\theta)]$ and the set of basic feasible solutions, $\mathbf{BFS} \equiv \{\boldsymbol{\theta} = (\alpha,\delta,\mu,\varepsilon)|\ \alpha,\delta\in [0, 1]$ and $\mu,\varepsilon\in [0, \infty)\}$ is the parameter constraints. Moreover, the ratio of mean informed trades to mean total trades is defined as

$$PIN \equiv \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} \tag{3}$$

However, in numeric computing, the problem (2) is subject to the basic feasible set $\mathbf{BFS}_{\mathrm{L_BT}} \equiv \{\boldsymbol{\theta} \in \mathbf{BFS}|\ L_B(\theta|\mathrm{T})$ do not lead to the FPE}, which is narrower than **BFS.** Namely, the poor expression of factored form $L_B(\boldsymbol{\theta}|\mathbf{T})$ may substantially distort the basic feasible set in computing and result in a biased PIN estimate.

2.1. Computing bias in estimating PIN and the remedy

Our argument regarding the adversarial effect of invalid expression of factored form echoes Lin and Ke (2011), who show that inaccurate expressions of $L_B(\boldsymbol{\theta}|\mathbf{T})$ result in underestimated PINs. For example, Easley et al. (2008) reformulate $L_B(\boldsymbol{\theta}|\mathbf{T})$ and propose $L_B(\boldsymbol{\theta}|\mathbf{T}) \equiv \sum_{i=0}^{I} L_I(\theta|B_i, S_i)$ as

$$L_I(\theta|B_i, S_i) \equiv \log[\alpha\delta\exp(-\mu)x^{B_i-M_i} + \alpha(1-\delta)\exp(-\mu)x^{S_i-M_i} + (1-\alpha)x^{B_i+S_i-M_i}] + (B_i + S_i)\log(\varepsilon + \mu) - 2\varepsilon + M_i\log(x), \tag{4}$$

where $M_i = \min(B_i, S_i) + \max(B_i, S_i)/2$, and $x = \varepsilon/(\mu + \varepsilon)$. Also the term $\log(S_i!B_i!)$ is dropped from (4). Lin and Ke (2011)[1] show, in a computing process, the MLE Problem (5) may generate the solutions different from those to Problem (2) due to FPE:

$$\max_{\theta\in \mathbf{BFS}_{\mathrm{L_IT}}} L_I(\theta|\mathbf{T}), \tag{5}$$

where $\mathbf{BFS}_{\mathrm{L_IT}} \equiv \{\boldsymbol{\theta} \in \mathbf{BFS}|\ L_B(\boldsymbol{\theta}|\mathbf{T})$ does not lead to an FPE}, which varies with $\mathbf{T}$ and the expression of $L_B(\boldsymbol{\theta}|\mathbf{T})$. Guaranteeing no overflow during the process of solving Problem (5), Lin and Ke also show that $\mathbf{BFS}_{\mathrm{L_IT}}$ is as follows:

[1] See the supplement to Lin and Ke (2011) in http://ssrn.com/abstract =1500828.

$$\mathbf{BFS}_{\mathrm{L_IT}} \equiv \left\{\theta \in \mathbf{BFS} \middle| \frac{\mu}{\varepsilon} < \exp\left(\frac{2E}{\max(B_{\max}, S_{\max})}\right) - 1\right\} \subset \mathbf{BFS}, \tag{6}$$

where $B_{\max}$ ($S_{\max}$) is the maximum of B_i (S_i) and E is approximately 710, which is the minimum of input values resulting in overflows for exponential function $\exp(\cdot)$ in the computer.

Assume that $\hat{\theta}$ is the solution to (5) and actual $\boldsymbol{\theta}$ belongs to $\mathbf{BFS}\backslash\mathbf{BFS}_{\mathrm{L_IT}}$. Accordingly, with (6), Lin and Ke (2011) argue that $\hat{\theta}$ always results in an underestimated PIN because an upper bound exists for μ/ε in $\mathbf{BFS}_{\mathrm{L_IT}}$. To eliminate FPE bias, they propose $L_B(\boldsymbol{\theta}|\mathbf{T}) \equiv \sum_{i=1}^{I} L_A(\theta|B_i, S_i)$ as

$$\begin{aligned} L_A(\theta|B_i, S_i) \equiv \log[\alpha\delta\exp(e_{1i} - e_{\max i}) + \alpha(1-\delta)\exp(e_{2i} - e_{\max i}) + (1-\alpha)\exp(e_{3i} - e_{\max i})] \\ + B_i\log(\varepsilon + \mu) + S_i\log(\varepsilon + \mu) - 2\varepsilon + e_{\max i}, \end{aligned} \tag{7}$$

where $e_{1i} = -\mu - B_i\log(1 + \mu/\varepsilon)$, $e_{2i} = -\mu - S_i\log(1 + \mu/\varepsilon)$, $e_{3i} = -B_i\log(1 + \mu/\varepsilon) - S_i\log(1 + \mu/\varepsilon)$, and $e_{\max i} = \max(e_{1i}, e_{2i}, e_{3i})$. Again, the term $\log(S_i!B_i!)$ is dropped from (7).

2.2. The likelihood function for scaled data

Let $\tilde{L}_I(\tilde{\theta}|\tilde{\mathbf{T}}) \equiv L_B(\boldsymbol{\theta}|\mathbf{T})$, where $\tilde{\mathbf{T}} \equiv \{(\tilde{B}_i, \tilde{S}_i)\}_{i=1}^{I} = \{(B_i/c, S_i/c)\}_{i=1}^{I}$ with $c \geq 1$ is the set of scaled data.[2] Then, using $\tilde{\mathbf{T}}$, we solve Problem (8) to obtain the estimate of $\tilde{\theta} \equiv (\tilde{\alpha},\tilde{\delta},\tilde{\mu},\tilde{\varepsilon})$:

$$\max_{\tilde{\theta}\in \mathbf{BFS}_{\tilde{\mathrm{L}}_I\mathrm{T}}} \tilde{L}_I(\tilde{\theta}|\tilde{\mathbf{T}}), \tag{8}$$

where $\mathbf{BFS}_{\tilde{\mathrm{L}}_I\mathrm{T}} \equiv \{\tilde{\theta} \in \mathbf{BFS}|\ L_B(\boldsymbol{\theta}|\mathbf{T}) \text{ does not lead to an FPE}\}$.

Based on Jackson (2013), $c\tilde{\mu} = \mu$ and $c\tilde{\varepsilon} = \varepsilon$. Then, we obtain $\mathbf{BFS}_{\tilde{\mathrm{L}}_I\mathrm{T}}$ analogous to $\mathbf{BFS}_{\mathrm{L_IT}}$ as follows:

$$\begin{aligned} \mathbf{BFS}_{\tilde{\mathrm{L}}_I\mathrm{T}} &\equiv \left\{\tilde{\theta} \in \mathbf{BFS} \middle| \frac{\tilde{\mu}}{\tilde{\varepsilon}} < \exp\left(\frac{2E}{\max(\tilde{B}_{\max}, \tilde{S}_{\max})}\right) - 1\right\} \\ &= \left\{\theta \in \mathbf{BFS} \middle| \frac{\mu}{\varepsilon} < \exp\left(\frac{2cE}{\max(B_{\max}, S_{\max})}\right) - 1\right\} \subset \mathbf{BFS}, \end{aligned} \tag{9}$$

where $\tilde{B}_{\max} = B_{\max}/c$ and $\tilde{S}_{\max} = S_{\max}/c$.

Similarly, with (9), the scaled data approach at best only partially lessen the adversarial effect of FPE because $\mathbf{BFS}_{\mathrm{L_IT}} \subset \mathbf{BFS}_{\tilde{\mathrm{L}}_I\mathrm{T}} \subset \mathbf{BFS}$. Moreover, $\mathbf{BFS}_{\tilde{\mathrm{L}}_I\mathrm{T}}$ varies with the function form of $\tilde{L}_I(\tilde{\theta}|\tilde{\mathbf{T}})$, and therefore the effectiveness of scaling data varies with the expression of likelihood.

3. Simulation analysis

We assign a random parameter vector $\boldsymbol{\theta} = (\alpha, \delta, \mu, \varepsilon)$, where $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and $\delta \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, to each of 2,500 hypothetical stocks and perform a simulation test. Specifically, for each pair of α and δ, we randomly generate 100 combinations of μ and ε for [600, 1,200] with a probability density function $f(\mu) = 1/600$ and for [0, 2,400] with a probability density function $f(\varepsilon) = 1/2400$.

For each hypothetical stock with $\boldsymbol{\theta}$, we simulate the number of *Buys* and *Sells* (B_i,S_i) for 60 trading days. Then, with the unscaled trade counts, we maximize both $L_A(\boldsymbol{\theta}|\mathbf{T})$ and $L_I(\boldsymbol{\theta}|\mathbf{T})$ in **BFS** using 125 different initial values proposed in Lin and Ke (2011).[3] Panels A and B of Figs. 1 show the estimates of μ/ε.[4]

Furthermore, given a sample $\mathbf{T} \equiv \{(B_i, S_i)\}_{i=1}^{60}$ of *Buys* and *Sells* counts, if $\max(B_{\max}, S_{\max}) > 1{,}000$, we adopt the following scale factor suggested by Jackson (2013):

$$\text{Recommended scale factor} = c = \max(B_{\max}, S_{\max})/1000.$$

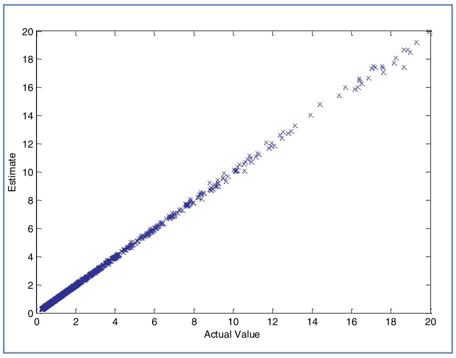
If $\max(B_{\max}, S_{\max}) \leq 1{,}000$, we set $c = 1$. Then, we calculate the scaled data set $\tilde{\mathbf{T}} \equiv \{(B_i/c, S_i/c)\}_{i=1}^{60}$ with $c \geq 1 = \{(\tilde{B}_i, \tilde{S}_i)\}_{i=1}^{60}$. With these scaled trade counts, we maximize both $L_A(\tilde{\theta}|\tilde{\mathbf{T}})$ and $L_I(\tilde{\theta}|\tilde{\mathbf{T}})$ in **BFS**. Panels C and D of Fig. 1 show the estimates of$\mu/\varepsilon(=\tilde{\mu}/\tilde{\varepsilon})$.

Panels A and C of Fig. 1 show that both unscaled and scaled data generate unbiased estimates of μ/ε using the accurate likelihood function. This result is consistent with the finding of Jackson (2013). However, Panels B and D show that the scaled data mitigate the impact of FPE but are not completely free from FPE bias in estimating PIN. A comparison of Panels B and D with Panels A and C suggests that the effect of scaled data varies with the inaccuracy of the likelihood expression for PIN estimation.

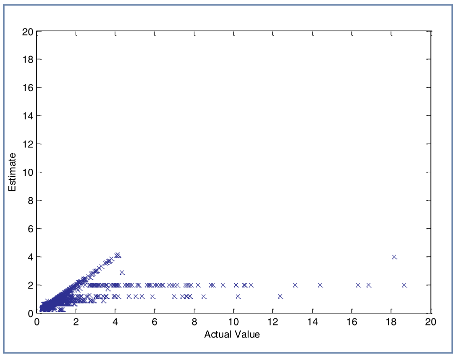
[2] $\mathrm{Log}(S_i!B_i!)$ is dropped, and thus we do not round the scaled trade counts to integers. Moreover, rounding the scaled counts to integers does not yield a significant difference in our results.

[3] Our estimation uses Matlab command "fminunc", which is effective for an unconstrained nonlinear optimization. The parameters α and δ are constrained using a logit conversion, whereas the other parameters, μ and ε, are constrained using a log conversion.

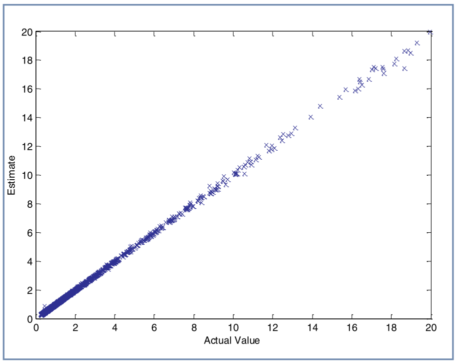
[4] The underestimation of PIN is due to the underestimation of μ/ε. We show these μ/ε estimates rather than PIN estimates because these estimates clearly show that the results are consistent with (6) and (9) in this paper.



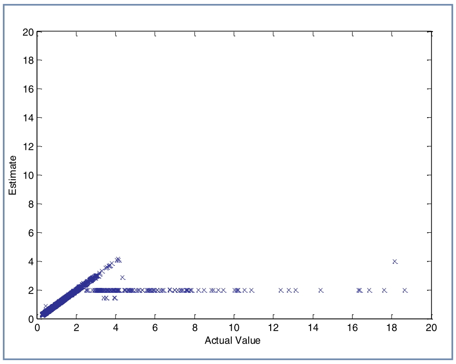
A. μ/ε estimates from L_A with unscaled trade counts



B. μ/ε estimates from L_I with unscaled trade counts



C. μ/ε estimates from L_A with scaled trade counts



D. μ/ε estimates from L_I with scaled trade counts

Fig. 1. Visual comparison of estimates and their actual values for μ/ε with the simulated sample
In Panels A and C, most of the μ/ε estimates using $L_A(\boldsymbol{\theta}|\mathbf{T})$ with unscaled or scaled trade counts lie along the 45-degree line. By contrast, in Panels B and D, most μ/ε estimates using $L_I(\boldsymbol{\theta}|\mathbf{T})$ with unscaled or scaled trade counts fall to the right of the 45-degree line and are systematically less than the actual values. Panel D shows that using scaled trade counts does not effectively mitigate the impact of FPE.

In sum, data scaling fails to effectively eliminate the FPE bias but rather mitigates its adversarial effect. The effectiveness of data scaling varies with the likelihood expression of the PIN model.

4. Suitable likelihood function for the data scaling approach

After adopting the scaled data (Jackson, 2013), there appears to be no concern that factorization may induce unintended biases in estimates. However, factorization is a key determinant of computing stability, and the effectiveness of scaling data depends on the factored Poisson pdf (in SAS or Matlab), which is not clearly specified in Jackson (2013).

Matlab, which is adopted in Jackson (2013), calculates the Poisson pdf $\mathrm{P}(x,\ \lambda)$ by its built-in function

$$\mathrm{poisspdf}(x, \lambda) = \exp(-\lambda + x\cdot\ln(\lambda) - \mathrm{gamma}\ln(x + 1)), \tag{10}$$

where gammaln(·) is the natural logarithm of the gamma function and usually is calculated by the approximation (Lanczos, 1964)

$$\mathrm{gamma}\ln(x + 1)\cdot$$
$$\approx \ln(x + \gamma + 1/2)\cdot(x + 1/2) - (x + \gamma + 1/2) + \ln(2\pi)/2 + \ln\left(c_0 + \sum_{n=1}^{14} \frac{c_n}{x + n}\right), \tag{11}$$

where the appropriate γ and c_i with $i = 0, 1, 2, \ldots, 14$ generate an error of less than 10^{-15} (see Passos, 2009). For Poisson distribution, gammaln$(x + 1)$ can be factored as $\sum_{n=1}^{x} \ln(n)$ for $x \geq 1$ and 0 for $x = 0$, which, however, may be time consuming.

The computing principles of Lin and Ke (2011) lend supports to an attentive selection of the expression in factored form. First, the factorization process of (10), namely, performing the multiplication by the exponential and logarithmic functions along with exponent rules is based on the first computing principle of Lin and Ke (2011). For example, to compute $e^x 3^y$ with $x = 1000$ and $y = -500/\ln(3)$, the expression $e^{x+y\ln(3)} = e^{500}$ is less vulnerable to the FPE than $e^x 3^y = e^{1000} 3^{-500/\ln(3)}$. The explanation is as follows. When e^{710} is the benchmark, computing $e^x 3^y$ leads to an overflow because $e^x = e^{1000}$ exceeds e^{710}, while computing $e^{x+y\ln(3)} = e^{500}$ does not. Second, for Easley et al. (1996) PIN model, if gammaln(·) is dropped during the computing process, Lin and

Kes (2011) second computing principle is needed for further factorization. The second principle requires that a large number be subtracted from the exponent of each exponential function in the likelihood factored by the first principle. For example, to compute $\log(e^x + e^y)$ with $x = 1800$ and $y = 900$, the expression $\log(e^{x-x} + e^{y-x}) + x = \log(e^0 + e^{-900}) + 900$ is more appropriate than $\log(e^x + e^y) = \log(e^{1000} + e^{900})$ because e^{1000} and e^{900} both lead to the overflows before log function is performed, while $e^0 + e^{-900}$ is greater than 1 and thus $\log(e^0 + e^{-900})$ would be calculated correctly. Accordingly, the factored PIN likelihood function proposed by Lin and Ke (2011) is robust for large argument values. In contrast, merely applying the PIN likelihood function with the factored (10) generates robust unbiased PIN estimates if and only if scaling data are adopted. Namely, a computing process via (10) is robust only for the Poisson pdf but not for the PIN joint pdf of (B_i, S_i).

The previous discussion shows that scaled data must be used along with the factored Poisson pdf of (10). With other factored likelihood functions, scaling data may still generate biased estimates. A more general remedy may still be Lin and Ke's (2011) two computing principles, which generate the suitable likelihood function for the approach proposed by Jackson (2013). If researchers attempt to construct the coding for Poisson pdf by themselves or when the adopted software fails to implement poisspdf(x, λ) or gammaln($x + 1$), the likelihood function of Lin and Ke is an effective choice for the data scaling approach.

5. Conclusions

Jackson (2013) suggests the scaled trade counts and Lin and Ke (2011) propose the factored likelihood function to estimate PIN for actively traded stocks. Jackson (2013) indicates that scaling data help avoid over/underflow problems resulting from large trades. However, we show that without the accurate (or suitable) likelihood expression this data scaling approach, despite being intuitive and easy to implement, still does not guarantee an unbiased PIN estimation.

Specifically, Jackson's (2013) recommended scale factor may not be suitable for all likelihood expressions. The data scaling method may be used along with (10), which is verified by Jackson. The likelihood function of Lin and Ke (2011) can be another choice.

Jackson (2013), however, does not specify his adopted likelihood expression and thus the readers who adopt a formula different from (10), such as $\exp(-\lambda)\,\lambda^x/x!$, may obtain unanticipated results. This note provides researchers with this reminder. Moreover, when the built-in Poisson pdf is not available, the factored likelihood function by Lin and Ke (2011) remains effective.

Based on our prior experience, with the likelihood function of Lin and Ke (2011), researchers may benefit from scaling data for the derivative-based optimization methods when these methods suffer from computing gradient for certain rare unexplained reasons. Similarly, the PIN likelihood function construed by (10) may be ineffective for computing gradient. Therefore, Jackson (2013) adopts a derivative-free method, which is not prevalently adopted due to its time-consuming nature.

Moreover, the choice of initial value appears to affect the performance of the two algorithms (e.g. Ersan and Alici, 2016; Gan et al., 2015). However, despite that adopting appropriate starting values the estimate may coverage quickly, it rarely solve FPE bias. Specifically, with FPE being present, only when initial values lie within the restricted feasible region, the estimate can converge to a solution, which is also necessary in the feasible region but may be biased. Namely, the combination of the scaled data and an excellent initial value determination may fail to solve FPE bias, either.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.frl.2018.12.017.

Reference

Akay, O., Cyree, K.B., Griffiths, M.D., Winters, D.B., 2012. What does PIN identify? Evidence from the T-bill market. J. Finan. Markets 15, 29–46.
Aslan, H., Easley, D., Hvidkjaer, S., O'Hara, M., 2011. The characteristics of informed trading: Implications for asset pricing. J. Empirical Finance 18, 782–801.
Duarte, J., Hu, E., Young, L., 2017. Does the PIN model mis-identify private information and if so, what is the alternative? Working Paper, available at SSRN: https://ssrn.com/abstract=2564369.
Duarte, J, Young, L, 2009. Why is PIN priced? J. Finan. Econ. 91, 119–138.
Easley, D., Engle, R.F., O'Hara, M., Wu, L., 2008. Time-varying arrival rates of informed and uninformed trades. J. Finan. Econ. 6, 171–207.
Easley, D., Lopez de Pardo, M.M., O'Hara, M., 2012. Flow toxicity and liquidity in a high frequency trading world. Rev. Financ. Studies 25, 1457–1493.
Easley, D., Kiefer, N.M., O'Hara, M., Paperman, J.B., 1996. Liquidity, information, and infrequently traded stocks. J. Finance 51, 1405–1436.
Easley, D., O'Hara, M., 1987. Price, trade size, and information in securities markets. J. Financ. Econ. 19, 69–90.
Ersan, O., Alici, A., 2016. An unbiased computation methodology for estimating the probability of informed trading. J. Int. Financ. Markets Instit. Money 43, 74–94.
Gan, Q., Wei, W.C., Johnstone, D., 2015. A faster estimation method for the probability of informed trading using hierarchical agglomerative clustering. Quantit. Finance 15 (11), 1805–1821.
Gan, Q., Wei, W.C., Johnstone, D., 2017. Does the probability of informed trading model fit empirical data? Financ. Rev. 52 (1), 5–35.
Jackson, D., 2013. Estimating PIN for firms with high levels of trading. J. Empirical Finance 24, 116–120.
Ke, W.-C., Lin, H.-W., 2017. An improved version of the volume-synchronized probability of informed trading (VPIN). Critical Finance Rev. 6 (2), 357–376.
Lanczos, C., 1964. A precision approximation of the gamma function. SIAM J. Numer. Anal. Series B 1, 86–96.
Lai, S., Ng, L., Zhang, B., 2014. Does PIN affect equity prices around the world? J. Financ. Econ. 114 (1), 178–195.
Lin, H.-W., Ke, W.-C., 2011. A computing bias in estimating the probability of informed trading. J. Financ. Markets 14, 625–640.
Passos, W.D., 2009. Numerical Methods, Algorithms and Tools in C#, first ed. CRC Press.
Preve, D., Tse, Y.-K., 2012. Estimation of time-varying adjusted probability of informed trading and probability of symmetric order-flow shock. J. Appl. Econ. 28, 1138–1152.