Journal of Risk **21**(2), 1–35
DOI: 10.21314/JOR.2018.399

Research Paper

BV–VPIN: Measuring the impact of order flow toxicity and liquidity on international equity markets

Rand Kwong Yew Low,[1] Te Li[2,3] and Terry Marsh[4,5]

[1]UQ Business School, University of Queensland, Brisbane, QLD 4072, Australia;
email: r.low@business.uq.edu.au

[2]Fu Foundation School of Engineering and Applied Science, Columbia University, New York, NY 10027, USA; email: tl2399@columbia.edu

[3]PF2 Securities Evaluations, 29th Floor, 85 Broad Street, New York, NY 10004, USA

[4]Quantal International Incorporated, Suite 1200, 455 Market Street, San Francisco, CA 94105, USA; email: terry.marsh@quantal.com

[5]Haas School of Business, University of California, 545 Student Services Building, Berkeley, CA 94720, USA

(Received November 6, 2017; revised February 13, 2018; accepted February 28, 2018)

ABSTRACT

Order flow toxicity is the measure of a trader's exposure to the risk that counterparties possess private information or other informational advantages. High levels of order flow toxicity can culminate in market makers providing liquidity at a loss or in the suboptimal execution of trades. From a regulatory perspective, high levels of toxicity can be harmful to overall market liquidity and precede precipitous drops in asset prices. The bulk volume–volume-synchronized probability of informed trading (BV–VPIN) Changes to sentence OK? model is one way of measuring the "toxicity" component of order flow, and it has been successfully applied in high-frequency trading environments. We apply the BV–VPIN to daily data from a range of international indexes in order to extend previous Change OK? analyses of its properties. We find that a rise in BV–VPIN effectively foreshadows high levels of volatility in the equity indexes of several countries. If a BV–VPIN futures

Corresponding author: R. K. Y. Low

Print ISSN 1465-1211 | Online ISSN 1755-2842

contract were to exist, we show that it would exhibit safe haven characteristics during market downturns. In particular, a simple active portfolio management strategy that times investments in equities (risk-free assets) when BV–VPIN levels are low (high) outperforms a buy-and-hold strategy. Thus, we find support for the application of BV–VPIN in international equity markets as a risk monitoring and management tool for portfolio managers and regulators.

Keywords: VPIN; bulk volume classification; international equities; liquidity; order flow toxicity.

1 INTRODUCTION

On May 6, 2010, the Dow Jones Industrial Average (DJIA) suffered a decline of 998.5 points. This event is the largest one-day point decline in the history of the DJIA and is regularly referred to as the "flash crash". Several explanations have been provided, eg, speculative trading by hedge funds (Patterson and Lauricella 2010; Phillips 2010), currency movements in the US dollar/Japanese yen exchange rate (Krasting 2010) and technical reporting difficulties on the New York Stock Exchange (NYSE) (Flood 2010). However, the most salient feature of the flash crash is the extreme lack of liquidity that occurred (Easley *et al* 2011b; CFTC–SEC 2010a,b). More specifically, Easley *et al* (2011b) attribute that tight liquidity to an unprecedented increase in order flow toxicity, a situation in which uninformed traders (eg, market makers) start to reduce their positions significantly to curtail the risk of significant losses due to adverse selection by informed traders (eg, hedge funds). Such actions are accompanied by market illiquidity and a significant fall in asset prices.

Easley *et al* (2012) develop the volume-synchronized probability of informed trading (VPIN) model as a measure of order flow toxicity, while Easley *et al* (2011b) show that the cumulative distribution function (CDF) of VPIN would have been able to predict the flash crash more than an hour before it happened. The VPIN is a metric of the probability that the liquidity provision process might fail and result in adverse price movements. Our work investigates the application of a version of VPIN – namely, bulk volume VPIN (BV–VPIN) – to daily international equities data in order to evaluate its effectiveness in ascertaining the probability of sharp market movements as well as its ability to act as a safe haven or hedging tool if implemented via a futures contract.

Comparing a range of trade classification algorithms that use tick data and the bulk volume of trades, Easley *et al* (2016) find that BV–VPIN is closely linked to information-based trading proxies such as hi–lo spreads and the permanent price effect of trades. BV–VPIN appears to be incrementally better than the original VPIN in discerning trade motivation from market data. Internationally, Abad and Yagüe

(2012) apply both BV–VPIN and the probability of informed trading (PIN) to the Spanish equities market and conclude that BV–VPIN is a useful proxy for adverse selection risk. They find in particular that the volume bucket size (VBS) input of the BV–VPIN model can be adjusted to capture transitory or permanent information in the trade data being analyzed. The authors further conclude that BV–VPIN has potential applications in the context of low-frequency analyses and should not be limited to purely high-frequency trading (HFT) environments.[1]

The BV–VPIN methodology has also been the subject of criticism. Andersen and Bondarenko (2014b) find results that contradict those of Easley *et al* (2011b) and document that BV–VPIN is a poor predictor of short-term volatility. They suggest that any predictive content is due to a mechanical relation with underlying trading intensity. However, Easley *et al* (2014) argue that the findings of Andersen and Bondarenko (2014b) are due to confusion in the methodology, analysis and conclusions drawn from their work. Further, Wu *et al* (2013) analyze the ninety-four most active futures contracts and find BV–VPIN to be a strong predictor of liquidity-induced volatility. Andersen and Bondarenko (2014a) and Andersen and Bondarenko (2015) continue to report that transaction-based classification schemes are more accurate than the BV strategies advocated by Easley *et al* (2011b) and Easley *et al* (2012).

Given the apparent importance of implementation and market context to the predictive content of BV–VPIN as an early warning signal of volatility, it is instructive to apply it to a variety of international equity markets using daily data. That is the main purpose of our paper. We analyze the impact of different values of the VBS and sample size applied as inputs in a BV–VPIN model based on the US market in order to ascertain the optimal criteria for application across all other countries in our data set. We report the threshold value that is optimal across a variety of countries. The economic value of BV–VPIN is evaluated by applying a scenario in which a fund manager exhibits a flight-to-quality action by alternating between investment in a risk-free security and in the stock market when the BV–VPIN is above or below

Changes to sentence OK?

[1] Given regulatory interest in promoting efficient market trading mechanisms, Easley *et al* (2011b) and Bell (2013) recommend that regulators use BV–VPIN as an early warning signal to herald the implementation of regulatory action to forestall crashes or identify unusual market conditions. Easley *et al* (2012) shows that BV–VPIN has forecasting power over volatility (toxicity-induced) and advocate its use as a valuable risk management tool for market-making activity. BV–VPIN has applications for trading strategies based on volatility arbitrage and for brokers who seek to ascertain the optimal time of execution (Easley *et al* 2015). Easley *et al* (2011a) detail the specifications of a BV–VPIN contract that could be used as a hedge against the risk of higher-than-expected levels of toxicity and to monitor such risk. VPIN has been applied in the empirical finance literature to evaluate the market reaction to public and private information (Vega 2006), market anomalies (Chen and Zhao 2012; Kang 2010), asset pricing (Aslan *et al* 2011) and more. Corcoran (2012) discuss how VPIN can be applied to detect mini-bubbles based on the idea of monitoring the probability of toxicity-induced liquidity crises.

the threshold value over the length of our data set. Our study includes the evaluation of the safe haven and hedge properties of BV–VPIN in the context of international equity markets.

We also show that BV–VPIN has potential applications beyond that of HFT. The CDF of BV–VPIN shows promise as a long-term predictor of market volatility. BV–VPIN also exhibits safe haven and hedge characteristics during well-known crisis periods, eg, the Asian financial crisis (AFC), the dot-com bubble (DCB), the great recession (GR) and the US credit rating downgrade (USCRD). Incorporating BV–VPIN as a flight-to-quality indicator in an asset management application results in the substantial outperformance of equity benchmark strategies for four out of the six countries analyzed in our study.

Our contribution is threefold. First, we apply the BV–VPIN to an international equities data set including the United States, the United Kingdom, Germany, Japan, Australia and China, and we analyze its performance over a data set spanning sixteen years.[2] Second, our data set allows us to determine the optimal thresholds for VPIN to predict large down movements across a range of international equity markets. Third, we analyze whether BV–VPIN would be useful with regard to its safe haven and hedging properties if it were implemented as a futures contract, as suggested by Easley *et al* (2011a). Our research objective is not to show that BV–VPIN works perfectly across all equity markets internationally. Instead, we investigate the equity markets in which it exhibits an adequate performance as a forward indicator of market turmoil in a low-frequency setting; this is as opposed to an HFT situation, which has already been heavily investigated in the literature.

Changes to sentence OK?

Changes to this sentence and the next OK?

The remainder of this paper is organized as follows. Section 2 presents the descriptive statistics of our data sets, while Section 3 details the BV–VPIN algorithm and parameters selected in our study. The results and conclusions of our paper are reported in Section 4 and Section 5, respectively.

2 DATA

Our data set includes daily equity returns and volume data from the US (Standard & Poor's 500 (S&P 500)), UK (Financial Times Stock Exchange 100 (FTSE 100)), German (Deutscher Aktienindex (DAX)), French (Cotation Assistée en Continu 40 (CAC 40)), Japanese (Nikkei 225), Chinese (Shanghai Composite Index (SHCOMP)) and Australian (AS30) indexes, sourced from the Bloomberg data terminal. The period covered is 1995–2015 for all countries except China. For China, our data set extends from 1997 to 2015.

Table 2 provides descriptive statistics for both our daily returns and volume data for all indexes. Japan is the only country to have a negative mean return. The standard

[2] A data set of futures contracts is applied by Wu *et al* (2013).

deviation of daily returns is similar across all countries, with Australia exhibiting the lowest. All indexes exhibit negative skewness. The US exhibits the highest levels of kurtosis, while Germany and China exhibit the lowest. The null hypothesis of normality is rejected at the 1% level for all countries using the Jarque–Bera test.[3] The daily volume characteristics are such that most countries exhibit similar mean daily volumes, with China (Germany) being the highest (lowest). The daily volume data for China also exhibits the highest positive skewness and kurtosis.

3 MEASURES OF ORDER FLOW TOXICITY

VPIN is related to order flow toxicity. The higher the order flow toxicity, the greater the potential for a negative impact on market makers as a result of being adversely selected by informed traders. When toxicity levels are too high, market makers will leave the market, resulting in a reduction in liquidity and short-term toxicity-induced volatility. Changes to sentence OK?

VPIN uses volume imbalance to signal the toxicity of order flow, and the overall level of volume determines the frequency for VPIN metric updates. VPIN is updated based upon fixed units of volume rather than fixed units of time. The most important feature of VPIN is its emphasis on volume as a critical variable in understanding price adjustments and its linkage to underlying information dynamics. In HFT markets, Easley *et al* (2012) argue that traders operate on a volume basis and focus on turning over holdings within specific numbers of contracts traded as opposed to over specific time intervals. We next explain the VPIN and BV–VPIN measures in detail and then analyze whether the latter's auxiliary information enables better prediction of adverse market conditions when applied to daily international index data.

3.1 VPIN

VPIN is based on the PIN model that Easley *et al* (1996) put forward as a measure of information asymmetry between informed and uninformed trades. PIN is premised on uninformed traders buying or selling regardless of whether new information exists as well as on informed trading occurring only when new information exists and buying (selling) upon arrival of good (bad) news. PIN is not a directly observable measure and is a function of the theoretical parameters of a microstructure model that is calculated using maximum likelihood estimation (MLE) by fitting a mixture of three Poisson distributions (Easley *et al* 1997). However, the numerical optimization procedure often has difficulty converging, and estimations may be biased (Abad and Yagüe 2012; Easley *et al* 2010; Lin and Ke 2011). These difficulties are exacerbated

[3] For brevity, we do not report the Jarque–Bera test statistic. These results can be provided upon request.

TABLE 1 Descriptive statistics of daily returns and trading volume of international indexes.

(a) Daily returns

Country	Index	Mean	SD	Skew	Kurt.	Min.	Max.	Sample size
United States	S&P 500	0.03	0.0121	−0.2449	8.0496	−9.47	10.96	5288
United Kingdom	FTSE 100	0.01	0.0117	−0.1639	5.9236	−9.27	9.38	5334
Germany	DAX	0.03	0.0150	−0.1273	4.1825	−8.87	10.80	5320
Japan	Nikkei 225	−0.0007	0.0153	−0.2901	5.3782	−12.11	13.23	5160
China	SHCOMP	0.02	0.0164	−0.2585	4.3751	−9.26	9.40	4355
Australia	AS30	0.02	0.0095	−0.5590	6.4182	−8.55	6.07	5315

(b) Daily volume

Country	Index	Mean	SD	Skew	Kurt.	Min.	Max.	Sample size
United States	S&P 500	9.39e+08	4.87e+08	0.5390	−0.2553	1.23e+06	2.96e+09	5288
United Kingdom	FTSE 100	9.98e+08	6.06e+08	0.6116	0.1695	5.00e+00	4.41e+09	5334
Germany	DAX	8.78e+07	6.25e+07	1.1347	3.3525	7.72e+05	4.97e+08	5320
Japan	Nikkei 225	9.48e+08	6.38e+08	1.0151	2.4677	7.09e+07	5.95e+09	5160
China	SHCOMP	7.75e+09	1.10e+10	3.1178	11.9300	1.23e+08	8.57e+10	4355
Australia	AS30	6.33e+08	4.34e+08	1.0159	3.5692	3.48e+05	5.64e+09	5315

The descriptive statistics of daily returns and volume data for indexes from six different countries. Our data set spans from January 1995 to December 2015, except for the SHCOMP (China), which spans from November 1997 to December 2015. The sample size indicates the number of days of data for each equities index. The mean, minimum and maximum are presented as percentages. SD denotes standard deviation.

by the sheer size of HFT data sets. PIN is based on a sequential trading model with Bayesian updates, as shown below:

$$\mathrm{PIN}_t = \frac{\alpha_t \mu}{\alpha_t \mu + 2\epsilon}. \tag{3.1}$$

The probability that new information will arrive within the time frame of the analysis is given by α, and μ is the arrival rate of informed traders, which are assumed to follow a Poisson process. The arrival rate of uninformed traders is given by ϵ. Easley *et al* (2008) propose a dynamic econometric microstructure model of trading and show that for a sufficiently large μ, the expected trade imbalance ($E[|V^{\mathrm{B}} - V^{\mathrm{S}}|]$) and the expected total number of trades ($E[|V^{\mathrm{B}} + V^{\mathrm{S}}|]$) are given by (3.2) and (3.3), respectively:

$$E[|V^{\mathrm{B}} - V^{\mathrm{S}}|] \approx \alpha\mu, \tag{3.2}$$

$$E[|V^{\mathrm{B}} + V^{\mathrm{S}}|] \approx \alpha_t \mu + 2\epsilon, \tag{3.3}$$

where V^{B} (V^{S}) is the volume traded against the ask (bid). While PIN uses an itemized classification scheme to distinguish between buy and sell volume, Easley *et al* (2012) propose an approach called "bulk classification" for VPIN. They argue that in an HFT environment, itemized approaches can be problematic. The authors suggest that a volume clock (instead of a calendar clock) should be applied with a fixed VBS. Using VBS is advantageous, as it is analogous to having a trading session split into periods of comparable information content and minimizes the impact of volatility clustering. Changes to sentence OK?

A trade classification algorithm is applied to identify the buy (V^{B}) and sell (V^{S}) volume. As all volume bars are of a fixed size, Easley *et al* (2012) show that V can be given by the following:

$$V = \frac{1}{N} \sum_{\tau=1}^{N} (V_\tau^{\mathrm{S}} + V_\tau^{\mathrm{B}}) = \alpha\mu + 2\epsilon. \tag{3.4}$$

For a given number of N volume bars, based on (3.1)–(3.4), it can be shown that VPIN is a good approximation of PIN and is given by

$$\mathrm{VPIN} = \frac{\sum_{\tau=1}^{N} |V_\tau^{\mathrm{S}} - V_\tau^{\mathrm{B}}|}{N \times V}. \tag{3.5}$$

A high VPIN value indicates large order imbalances that may threaten liquidity provisions, hence its application as a market liquidity indicator. With access to tick data or order book information, VPIN can be operationalized by adding the buy and sell volume of each volume bar. However, tick data and market order data are often inaccurate, incomplete or expensive to access. Thus, trade classification algorithms are commonly used to approximate the buy or sell volume.

3.2 BV–VPIN

VPIN is often applied with one of three trade classification algorithms: tick rule VPIN (TR–VPIN), BV–VPIN and Lee–Ready VPIN (LR–VPIN). Both TR–VPIN and BV–VPIN are level-one algorithms that only use trade price data. LR–VPIN is a level-two algorithm that uses both trade and quote data. It is commonly applied in markets where distinguishing the initiating side of the trade is difficult.[4] Changes to sentence OK?

Chakrabarty *et al* (2012) report that BV–VPIN is a more time-efficient methodology, whereas TR–VPIN produces more accurate measures of order imbalance and order flow toxicity. The authors find that both TR–VPIN and BV–VPIN perform well when identifying the aggressor side of trading, with TR–VPIN being slightly more accurate. Easley *et al* (2016) examines the accuracy and efficiency of BV–VPIN, LR–VPIN and the aggregated tick rule. They conclude that tick rule approaches are relatively good classifiers of aggressor trades. However, BV–VPIN is shown not only to be sufficiently accurate for classifying buy and sell trades but also to go beyond tick-based approaches by providing insight into other proxies for underlying trade information. Moreover, it has the advantage of requiring substantially less data for implementation.

In our study involving the analysis of daily data, we apply bulk volume classification to calculate the probability (ie, portion) of buy and sell in each volume bar. Trades are aggregated over volume intervals, and the price change between the two consecutive intervals is used to approximate the percentage of buy and sell order flow. The amount of volume that is classified as buy ($\hat{V}_\tau^{\mathrm{B}}$) or sell ($\hat{V}_\tau^{\mathrm{S}}$) is given by (3.6) and (3.7), respectively:

Later, I am reasonably confident that you used 'df' to denote the 'degrees of freedom' (see Table 2, for example), and have therefore made it 'df' in keeping with journal style. But I am not sure whether that is what is meant by df here, or whether this is the multiplication of two variables, or perhaps even $\mathrm{d}f$, where 'd' is the differential element. Please check throughout and mark any that need to be changed.

$$\hat{V}_\tau^{\mathrm{B}} = V_\tau t\left(\frac{P_\tau - P_{\tau-1}}{\sigma_{\delta p}}, df\right), \tag{3.6}$$

$$\hat{V}_\tau^{\mathrm{S}} = V_\tau\left[1 - t\left(\frac{P_\tau - P_{\tau-1}}{\sigma_{\delta p}}, df\right)\right], \tag{3.7}$$

where P_τ is the single price ascribed to each current volume bar and $P_{\tau-1}$ is the price of the previous volume bar. σ_{δ_p} represents the volume-weighted standard deviation of price change over two consecutive volume bars. A volume bar is denoted by τ. A Student t distribution is used to approximate the price changes, and the probability of buy and sell volume is calculated using the CDF of the empirical price change distribution. The underlying trade is more likely to be buyer-initiated (seller-initiated) if the price change is positive (negative) as well as larger in magnitude relative to the

[4] Baker and Kiymaz (2013) state that order flow is defined as the number of trades in which the buyer was the aggressor minus the number of trades in which the seller was the aggressor. In a limit-order market (over-the-counter market), the aggressor is the agent placing a market order (requesting a quote).

distribution of past price changes. Volume-weighted standard deviation is calculated using

$$\sigma_{\triangle P_i} = \sqrt{\frac{\sum_{\tau=1}^{n} V_{i,\tau}(\triangle P_{i,\tau} - \overline{\triangle P_i})^2}{\sum_{\tau=1}^{n} V_{i,\tau}}}. \tag{3.8}$$

4 RESULTS

Easley *et al* (2016) show that the choices of VBS and sample length greatly impact the implementation and estimation of BV–VPIN, and that the suitable selection of these variables is dependent on optimizations being performed and should reflect the nature of trades within the market investigated. In Section 4.1, we discuss the choices available for different parameters of BV–VPIN and the values selected in our study across different markets.

To protect against the risks of adverse selection and preserve the integrity of the liquidity provision process, two approaches have been proposed. First, Zweig (2012) reports that exchanges can dynamically adjust the speed of the trading engine. If bids are being struck at such a speed that market makers are unable to adequately fill liquidity demands, they will be forced out, resulting in a liquidity crash. Thus, the trading engine would decelerate (accelerate) the speed of matches occurring at the bid (ask) price. In such an approach, BV–VPIN can be used as an early-warning indicator of when exchanges should take action, and super-computing resources may be applied accordingly (Wu *et al* 2013). By calculating BV–VPIN for multiple international markets, we analyze VPIN's ability to forecast sharp downward movements at different thresholds. We take the view of a global-macro fund manager who decides to invest in each country's equity index when risks are low and switches to the risk-free rate when risks are high. Henceforth, the economic impact of VPIN is evaluated by switching to the risk-free rate (ie, US one-month treasury bills (T-bills)) when VPIN levels are high, and investing in the equity index when VPIN levels are low. We report the outcomes of this analysis in Section 4.2.

Second, Easley *et al* (2011a) outline specifications for a VPIN futures contract that offers market makers a hedge against the risk of higher-than-expected levels of toxicity (ie, protection against a rise in the probability of adverse selection for liquidity providers). Market makers purchase protection when their inventories have risen beyond normal levels, and informed traders sell that protection once their orders have been fulfilled, thus allowing them to monetize their private knowledge that their contribution to toxicity has ceased. In Section 4.3, we analyze the safe haven and hedge properties of VPIN if implemented as a futures contract in accordance with Easley *et al* (2011a). Using the methodology of Baur and McDermott (2010) and Low *et al* (2016), we analyze whether the VPIN of each country exhibits safe haven and hedging properties against each country's equity index during the extreme quintiles

Changes to sentence OK?

Change OK?

TABLE 2 Parameters of BV–VPIN model.

Description	Parameters	Selected parameters
Number of volume bars	1000, 2000, 3000, 4000	2000
Rolling window size (bars)	10, 20, 50, 100	50
Distribution for price changes between volume bars	Normal, Student t (df $=$ 0.1, 0.25, 1, 5, 10)	Student t (df $=$ 5)
Student t distribution with df $=$ 5	Normal, lognormal	Lognormal

The key parameters of the BV–VPIN model when applied to daily data. The values used in our analysis are given in the third column.

of the equity returns distribution (eg, 10%, 5%, 1%) and well-known crisis periods.

4.1 Parameter selection for BV–VPIN

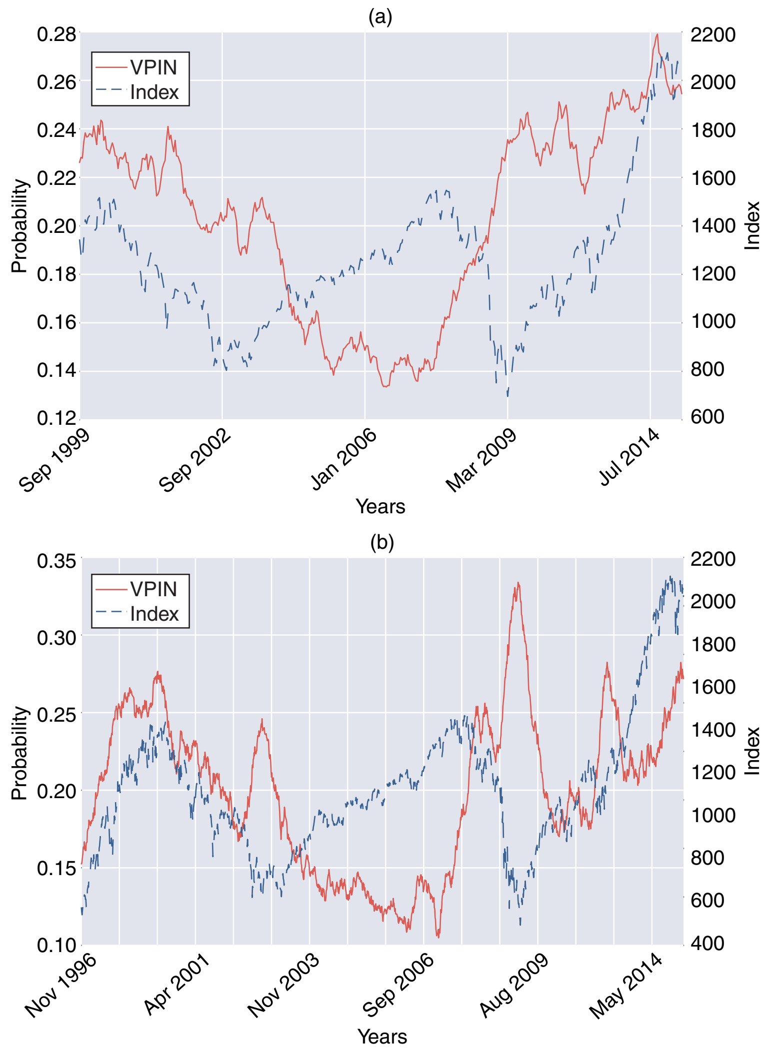
BV–VPIN takes price and volume as input parameters. Table 2 lists the key parameters to choose when calculating BV–VPIN. We investigate the optimal parameters based on the US data set and apply the same parameters to all other countries throughout our exploratory study. The US equities market continues to be one of the largest and most heavily traded markets globally, and therefore it is an appropriate choice for a test case.

The bulk volume classification algorithm requires each volume bar to be associated with a single price. Wu *et al* (2013) examine the performance under different pricing methodologies (eg, average, weighted average, median and weighted median) and find that the results do not differ significantly. We follow Easley *et al* (2016), who suggest using the last price of each volume bar to represent the price of the bar.

The prior literature applies fixed values (eg, 1000, 5000) to each volume bar. Our study calculates BV–VPIN across equity indexes from six countries, where each market has different trading volumes. Therefore, it is impractical to use a single value for the volume bar in each different market. Thus, we fix the total number of bars such that the volume of each bar is approximately equal to the total volume divided by the number of bars. Change OK?

Figure 1 shows the BV–VPIN of the US market with different volume sizes, keeping all other parameters fixed. Small volume sizes produce a more volatile BV–VPIN, increasing the degree of noise in the BV–VPIN's CDF. Applying larger volume sizes "smoothens" the BV–VPIN but may result in overfitting the BV–VPIN such that large impending price movements in the underlying equities index go undetected. For each country investigated, the optimal choice for the volume bar size may differ. As can be seen in Figure 1, a selection of 200 for the total number of volume bars looks to Change OK?

FIGURE 1 BV–VPIN and the CDF of BV–VPIN calculated for the US market (S&P 500) with different volume bar sizes. [Figure continues on next page.]

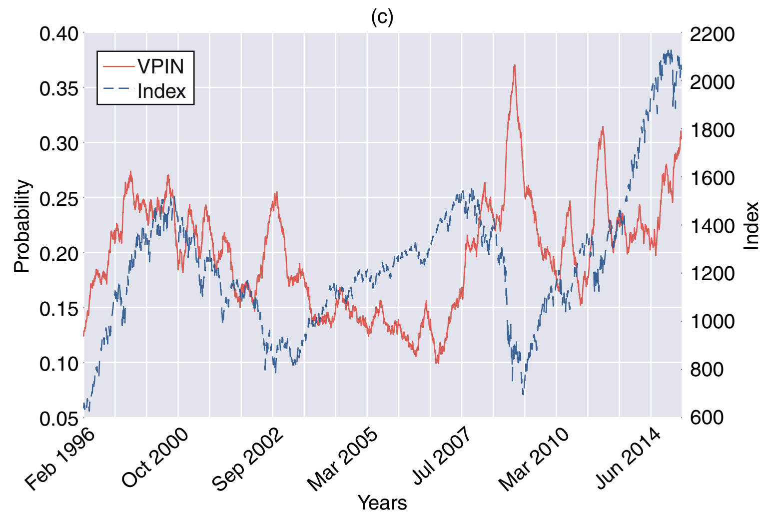


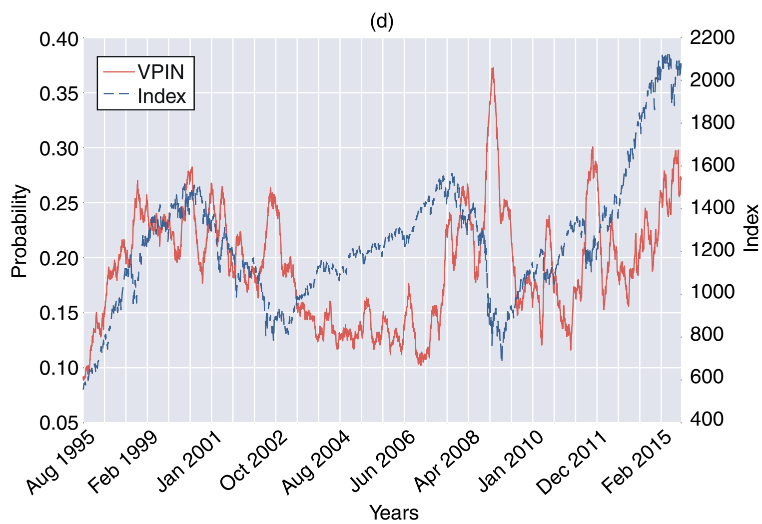
(a) 1000 volume bars. (b) 2000 volume bars.

have roughly the right balance of capturing salient downward features in the S&P 500 index.

The rolling window size (or VBS) is another important input parameter. Figure 2 shows the marginal effect on the BV–VPIN curve of different rolling window sizes.

FIGURE 1 Continued

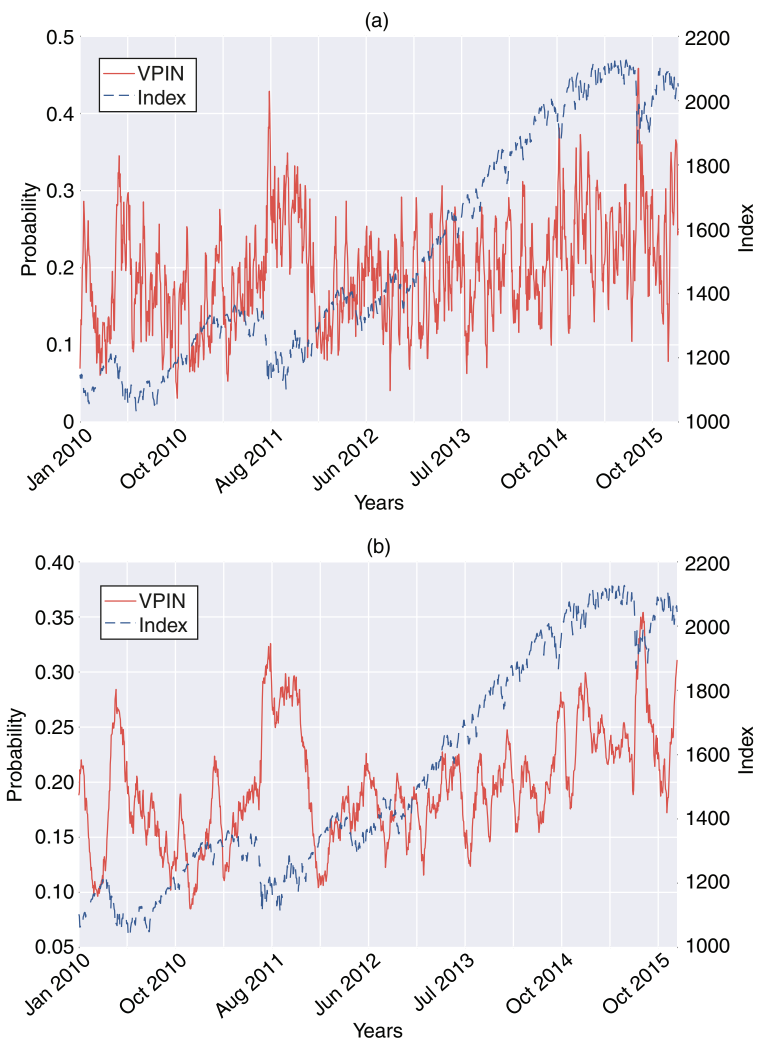




(c) 3000 volume bars. (d) 4000 volume bars.

Smaller window sizes follow price movements more closely but exhibit higher volatility. Larger window sizes result in less volatile and smaller BV–VPIN peaks. As our application of BV–VPIN is in the context of a long-term indicator of market liquidity in a low-frequency environment, we select 50 as the window size to ensure that large adverse price movements are detected.

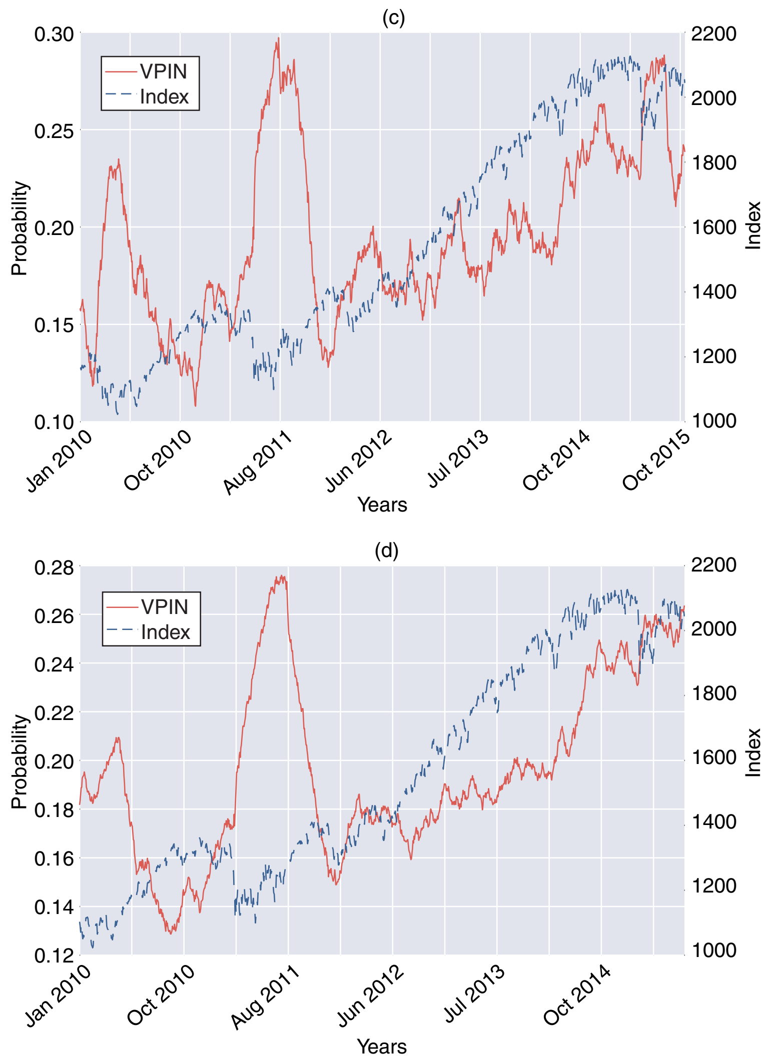
FIGURE 2 BV–VPIN and the CDF of BV–VPIN calculated for the US market (S&P 500), where rolling window lengths vary. [Figure continues on next page.]



(a) Rolling size = 5. (b) Rolling size = 20.

BV–VPIN classifies trade volume based on the distribution of price changes over volume bars. Easley *et al* (2012) suggest that for high-frequency data, a suitable choice is a Student t distribution with a d f of 0.25. For daily data, we have fitted our distributions using both normal and Student t distributions. Figure 3(a) shows the

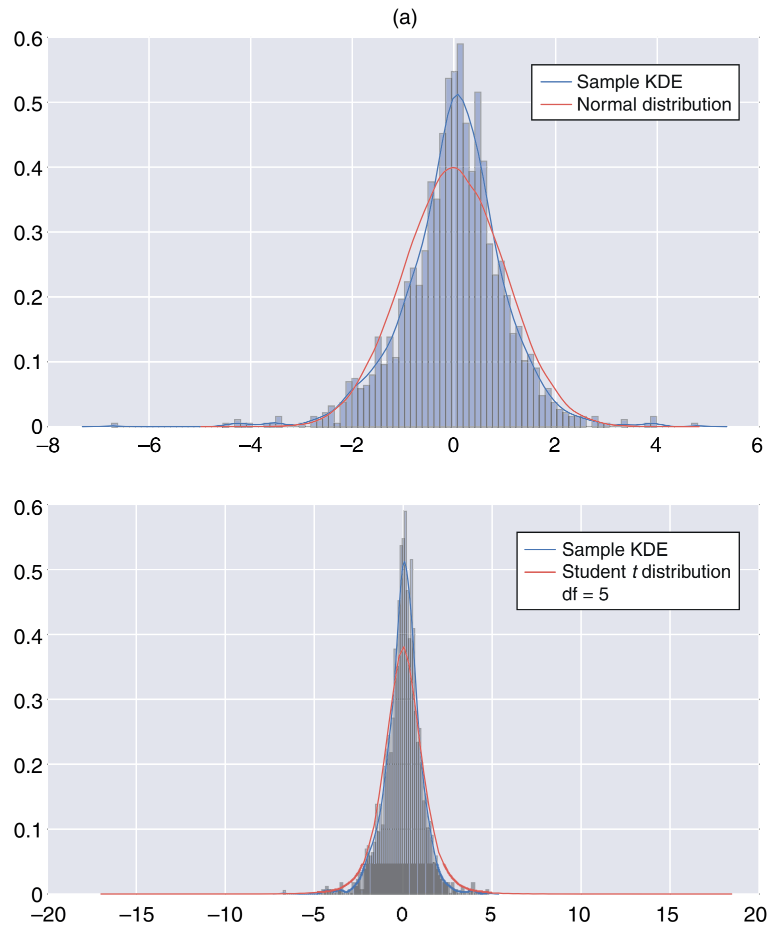
FIGURE 2 Continued.



(c) Rolling size = 50. (d) Rolling size = 100.

kernel density of price changes (KDE) versus normal and Student t distributions. It is intuitive to expect the Student t distribution will provide a superior fit to the normal distribution due to the prevalence of kurtosis within our data set of international equity indexes. Upon applying both the normal and Student t distributions, our results show

FIGURE 3 Distribution fitting of price changes (across volume bars) of BV–VPIN for the US market (S&P 500). [Figure continues on next page.]

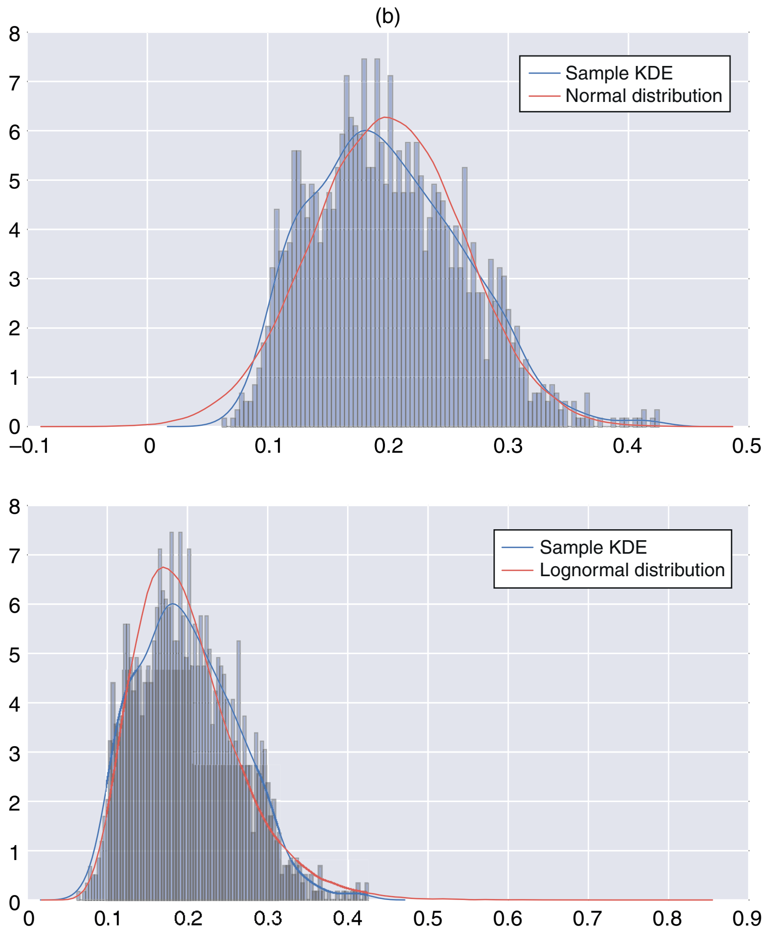


(a) Distribution of price change.

no obvious differences between the two. We select the Student t distribution with $\mathrm{df} = 5$.

Applying BV–VPIN alone is not an effective early warning signal. Easley *et al* (2011b) shows that the BV–VPIN rises almost simultaneously during prices drops and reaches its peak when prices have stabilized. Easley *et al* (2014) shows that the CDF of BV–VPIN reaches its peak before BV–VPIN and successfully predicts the impending flash crash in the e-mini S&P 500. As the BV–VPIN is calculated using

FIGURE 3 Continued.



(b) Distribution of VPIN.

absolute volume balances, we apply the lognormal distribution to approximate the VPIN distribution. Figure 3(b) shows that the sample distribution of the S&P 500 VPIN is more likely to follow a lognormal distribution. Wu *et al* (2013) suggest that the lognormal distribution accurately describes the BV–VPIN sample. They also recommend using the truncated lognormal distribution to avoid the occurrence of extremely small BV–VPIN values. For daily data, as BV–VPIN values are significantly larger than zero, we use the ordinary lognormal distribution with a CDF given by

$$\frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\ln x - \mu}{\sigma\sqrt{2}}\right)\right], \tag{4.1}$$

where erf is the Gaussian error function, μ is the average of the log of BV–VPIN values, and σ is the standard deviation.

4.2 BV–VPIN as an early warning signal

Figure 4 presents the BV–VPIN and CDF of BV–VPIN for equity indexes in different countries. We define a "VPIN event" as one that occurs when the CDF of BV–VPIN exceeds a threshold. Given that BV–VPIN follows a lognormal distribution, a threshold may be selected based on mean and variation.[5] In our implementation, we select 0.6, 0.8 and 0.9 as intuitive threshold values.

BV–VPIN tends to stabilize at 0.2 and then rise in concert with large price movements in the underlying index. For example, in Figure 4(a), the all-time high occurs during 2008–9, which is also the most volatile period for the S&P 500. It is notable across all countries that BV–VPIN is negatively correlated with level movements in the underlying index. However, BV–VPIN increases almost simultaneously when the index starts to fall.

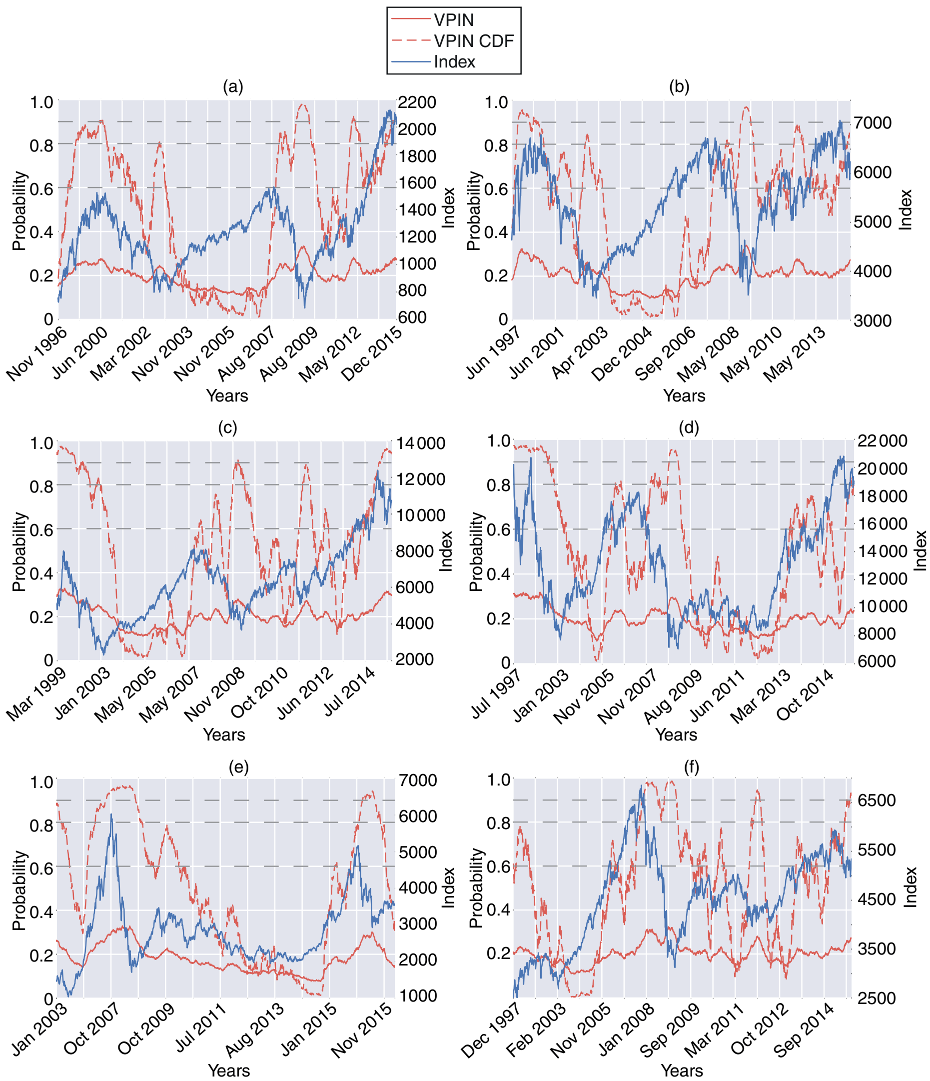
The CDF of BV–VPIN rises earlier and quicker than BV–VPIN and is more effective as an early-warning signal. All markets exhibit high BV–VPIN levels during well-known crises (eg, 2008 GR, 2011 USCRD). In the later months of our data set, we can see that BV–VPIN is relatively high, indicating that markets remain susceptible to volatility and liquidity shocks despite the best efforts of financial regulators and central banks to improve credit risk and market risk exposures within the financial system.

As BV–VPIN is proportional to the absolute difference between buy and sell volumes, any sudden and large upward price movements should result in BV–VPIN rising sharply. However, empirically, we are more likely to observe large and sudden price drops across US and international equities (Ang and Chen 2002; Longin and Solnik 1995; Low *et al* 2013). Thus, our analysis of prediction accuracy investigates the effectiveness of the CDF of BV–VPIN in forecasting impending index falls. Change OK?

Table 3 shows the prediction accuracy with different thresholds applied. Using 0.9 as our threshold value, we have a 100% prediction accuracy in four markets including the United States, the United Kingdom, China and Australia, and a 75% accuracy in Germany. For the same threshold, only one BV–VPIN event is observed in Japan. Using a threshold value of 0.8, we have a 75% prediction accuracy in only two countries. Moreover, our prediction accuracy is less than 50% in all countries if

[5] Wu *et al* (2013) use two standard deviations greater than the average value to denote a VPIN event.

FIGURE 4 BV–VPIN and CDF of BV–VPIN for international equities indexes.



(a) United States. (b) United Kingdom. (c) Germany. (d) Japan. (e) China. (f) Australia.

we use 0.6 as our threshold. It is intuitive to expect that lower thresholds will increase the number of false alarms, whereas a higher threshold might miss market downturns. The results presented in Table 3 are based on Figure 4. As BV–VPIN can hover around the threshold values, we only consider the first breach of a threshold value to be a Changes to sentence OK?

TABLE 3 Prediction properties of BV–VPIN. [Table continues on next page.]

(a) United States

Threshold	No. of BV–VPIN events	No. of price drops	Prediction accuracy (%)
0.9	4	4	100
0.8	5	3	60
0.6	5	2	40

(b) United Kingdom

Threshold	No. of BV–VPIN events	No. of price drops	Prediction accuracy (%)
0.9	3	3	100
0.8	4	3	75
0.6	14	6	43

(c) Germany

Threshold	No. of BV–VPIN events	No. of price drops	Prediction accuracy (%)
0.9	4	3	75
0.8	5	3	60
0.6	8	3	38

BV–VPIN event.

Intuitively, when a BV–VPIN event occurs, it is important to evaluate both the magnitude of the fall in the underlying index and the time horizon over which it occurs. This allows the portfolio manager to understand how much time they have to exit all of their equity positions if they decide to heed the BV–VPIN signal. Table 4 presents both the prediction accuracy and the index downturn with respect to different time horizons. We report our results, applying a threshold value of 0.9 for all countries except Japan, for which a threshold value of 0.8 is used. We record the dates associated with each BV–VPIN event and use the index level at these dates as the benchmark level. We observe prices 5, 20, 60 and 120 trading days after the BV–VPIN event. Based on these values, we calculate the fall in the index for each BV–VPIN event over different time horizons.

TABLE 3 Continued.

(d) Japan

Threshold	No. of BV–VPIN events	No. of price drops	Prediction accuracy (%)
0.9	1	0	0
0.8	4	3	75
0.6	10	3	30

(e) China

Threshold	No. of BV–VPIN events	No. of price drops	Prediction accuracy (%)
0.9	2	2	100
0.8	1	2	50
0.6	3	1	33

(f) Australia

Threshold	No. of BV–VPIN events	No. of price drops	Prediction accuracy (%)
0.9	3	3	100
0.8	4	1	25
0.6	7	1	14

The prediction properties of BV–VPIN. A BV–VPIN event occurs when the CDF of BV–VPIN exceeds the threshold value. Price drops are ascertained by any negative return that occurs within a horizon of thirty to sixty days after the BV–VPIN event occurs. Prediction accuracy is determined as the number of actual negative returns on the respective index divided by the total number of BV–VPIN events.

The calculation for the average index fall for negative returns is based solely on BV–VPIN events that resulted in an index fall. For example, over the one-month prediction horizon for the US index, there were four BV–VPIN events, two of which resulted in index falls and two that were false alarms. The calculation for the average price drop for all returns is based on all BV–VPIN events regardless of whether there was an index drop after the event. This allows us to understand the average index drop if BV–VPIN thresholds are breached while taking into account the impact of false

TABLE 4 Prediction horizon of BV–VPIN. [Table continues on next page.]

(a) United States

Prediction horizon (days)	Prediction accuracy	Average price drop (negative returns)	Average price drop (all returns)
5	3/4	−3.83%	−1.09%
20	3/4	−7.00%	−3.27%
60	2/4	−5.42%	2.74%
120	2/3	−13.06%	−4.60%

(b) United Kingdom

Prediction horizon (days)	Prediction accuracy	Average price drop (negative returns)	Average price drop (all returns)
5	2/3	−3.04%	4.03%
20	2/3	−6.81%	−2.46%
60	1/3	−18.41%	−4.72%
120	2/3	−2.70%	−1.25%

(c) Germany

Prediction horizon (days)	Prediction accuracy	Average price drop (negative returns)	Average price drop (all returns)
5	1/3	−8.55%	−0.16%
20	1/3	−1.12%	2.15%
60	2/3	−4.43%	−2.26%
120	3/3	−3.33%	−3.33%

alarms. The average index drop after a BV–VPIN event is calculated as follows:

$$\mathrm{PD}_{\mathrm{avg},t_f} = \frac{1}{t_f}\sum_{i=1}^{t_f} R_i, \tag{4.2}$$

$$R_i = \frac{P_i - P_{t_0}}{P_{t_0}}. \tag{4.3}$$

The day the BV–VPIN event occurs is given by t_0, and t_f represents five, twenty, 60 or 120 trading days after the BV–VPIN event. When calculating the average price drop

TABLE 4 Continued.

(d) Japan

Prediction horizon (days)	Prediction accuracy	Average price drop (negative returns)	Average price drop (all returns)
5	3/4	−4.33%	−3.05%
20	3/4	−4.87%	−2.10%
60	2/3	−9.44%	−5.82%
120	2/3	−9.48%	−5.86%

(e) China

Prediction horizon (days)	Prediction accuracy	Average price drop (negative returns)	Average price drop (all returns)
5	0/2	N/A	1.62%
20	1/2	−4.05%	4.88%
60	1/2	−19.15%	−0.06%
120	2/2	−5.57%	−5.57%

(f) Australia

Prediction horizon (days)	Prediction accuracy	Average price drop (negative returns)	Average price drop (all returns)
5	3/4	−3.60%	−2.44%
20	3/4	−2.35%	−1.51%
60	2/3	−16.04%	−10.67%
120	2/3	−20.20%	−13.26%

The prediction properties of BV–VPIN when using fixed thresholds of 0.9 for all countries, except Japan. For Japan, a threshold value of 0.8 is used. Prediction accuracy denotes the number of index drops out of the total number of BV–VPIN events in which the CDF of BV–VPIN exceeds the threshold value. The average index change indicates the average index returns based on the number of days after a BV–VPIN event occurs. The average index change (negative) indicates the average index returns only when the BV–VPIN event precedes a drop in the index.

for negative returns, we only include $R_i < 0$. In this scenario, the average price drop is based solely on the average of the two BV–VPIN events that resulted in price drops. In this manner, we understand the magnitude of the price drop when the BV–VPIN threshold is breached and a price drop has been accurately predicted.

Table 4 shows that prediction accuracy and average price drops vary across different markets and prediction horizons. We see that when the BV–VPIN event accurately predicts a price drop, we can expect to see a negative return of more than 9% in the following period of 60 or 120 days (except for Germany). However, within the first five days, we observe price drops of about 5% or less (except for China and Germany). One can conclude that once the BV–VPIN threshold is breached, a portfolio manager has at least thirty days to exit their positions to minimize the magnitude of losses. When accounting for all BV–VPIN events (ie, including the impact of false alarms), we see that after 120 trading days, all six countries experienced drops in the index, ranging from −1.25% (United Kingdom) to −13.26% (Australia). Within the first five and twenty trading days, we observe an index drop in four countries. Our results indicate that there are substantial drops in each country's index after a BV–VPIN event.

4.3 BV–VPIN as a hedge and safe haven investment

To evaluate the hedge and safe haven properties of BV–VPIN if implemented as a futures contract, our analysis utilizes the following model, as shown in (4.4)-(4.6):

$$r_{\text{VPIN},t} = a + b_t r_{\text{index},t} + \varepsilon_t, \tag{4.4}$$

$$b_t = c_0 + c_1 D(r_{\text{index}} q_{10}) + c_2 D(r_{\text{index}} q_5) + c_3 D(r_{\text{index}} q_1), \tag{4.5}$$

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}. \tag{4.6}$$

Equation (4.4) models the relationship between the CDF of BV–VPIN ($r_{\text{VPIN},t}$) and each individual country's equity index ($r_{\text{index},t}$). The parameter b_t is modeled as a dynamic process given by (4.5). Indicator variables ($D(\cdots)$) are applied to capture extreme market movements, taking a value of one if the stock market return at time t exceeds a certain threshold given by the 10%, 5% and 1% quantile of the return distribution, and zero otherwise. The residual term ε_t is modeled as a Glosten, Jagannathan and Runkle generalized autoregressive conditional heteroscedasticity (GJR–GARCH) process (Glosten *et al* 1993) that takes into account volatility clustering and asymmetries in the autoregressive conditional heteroscedasticity (ARCH) process. All equations are jointly estimated with MLE. These models have been applied in analyzing the safe haven and hedge properties of commodities such as precious metals (Baur and McDermott 2010) and diamonds (Low *et al* 2016). Change OK?

If any of the parameters (c_1, c_2, c_3) are significantly different from zero, this suggests a relationship between BV–VPIN and the country's index. If the parameters in (4.5) are nonpositive, BV–VPIN acts as a weak safe haven. If the parameters are negative and statistically different from zero, BV–VPIN becomes a strong safe haven. When c_0 is zero or negative and the sum of the parameters c_1 to c_3 are not jointly positive (exceeding the value of c_0), BV–VPIN serves as a hedge, where negative c_0 suggests strong hedging attributes, while a value of zero indicates weak hedging attributes.

TABLE 5 Hedge and safe haven properties of BV–VPIN.

				Safe haven quantiles					
Country	**Index**	**Hedge**	***t*-stats**	**10%**	***t*-stats**	**5%**	***t*-stats**	**1%**	***t*-stats**
US	S&P 500	0.055	3.373	−0.178***	−3.846	0.033	0.697	−0.041	−1.187
UK	FTSE 100	0.060	3.566	−0.191***	−4.160	0.104	2.263	−0.144***	−3.763
Germany	DAX	0.051	4.248	−0.155***	−4.188	0.012	0.319	−0.061**	−2.003
Japan	Nikkei 225	0.041	2.971	−0.133***	−3.930	−0.003	−0.010	−0.062**	−2.261
China	SHCOMP	0.144	8.356	−0.286***	−5.542	−0.021	−0.382	−0.017	−0.408
Australia	AS30	0.090	3.690	−0.207***	−3.456	−0.130**	−1.967	−0.018	−0.341

The estimation results for the coefficients in (4.5). Negative coefficients in the hedge column (c_0) signify that the asset is a hedge against the index. Zero (negative) coefficients in extreme market conditions, namely quantiles 10% (c_1), 5% (c_2) and 1% (c_3), indicate that the asset is a weak (strong) safe haven. Each column of t-statistics is associated with the coefficient column to the left as an indication of the significance level. *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

The estimated results of the regression model given by (4.4)–(4.6) are reported in Table 5, which contains the estimates of c_0, c_1, c_2 and c_3. Under the GJR–GARCH model, it is shown that BV–VPIN is a strong safe haven in all six countries within the 10% quantile, with strong statistical significance. At the 1% quantile, BV–VPIN remains an effective safe haven in all six countries, particularly in the United Kingdom, Germany and Japan, where the coefficients are statistically significant. Only Japan, China and Australia exhibit safe haven qualities at the 5% level, with Australia's results being statistically significant. All six countries have positive coefficients for c_0, suggesting that BV–VPIN does not exhibit suitable qualities for hedging.

Equation 4.7 allows the analysis of hedge and safe haven properties during crisis subsample periods. By identifying well-known financial crisis periods, we apply time dummies equal to one if the returns fall within the predefined period, and zero otherwise. We define starting dates and assume that most of effects of each crisis occurred in the first twenty trading days (approximately one month) following its start date (Baur and McDermott 2010). Four major financial events occurred in the sample period of our investigation, namely the AFC (October 22, 1997 to November 19, 1997), DCB (March 10, 2000 to April 7, 2000), GR (September 12, 2008 to October 2, 2008) and USCRD (July 23, 2011 to August 12, 2011): Changes to sentence OK?

$$b_t = c_0 + c_1 D(\text{AFC}) + c_2 D(\text{DCB}) + c_3 D(\text{GR}) + c_4 D(\text{USCRD}). \tag{4.7}$$

If the parameters c_1, c_2, c_3 or c_4 are zero or negative, BV–VPIN is a safe haven during the respective crisis period. Alternatively, a positive parameter means that the asset moves in tandem with the market and is not a safe haven.

Table 6 presents the estimation results of the model in (4.7). For Germany, China and Australia, the results for the AFC and DCB are unavailable due to the sample length of our data set and the structure of our empirical analysis applied to calculate BV–VPIN. When calculating BV–VPIN, we group trade data across several days into one volume bar, then compute BV–VPIN based on a rolling window of fifty volume bars. Thus, there are no BV–VPIN values for the initial forty-nine volume bars. Moreover, the trading volume is relatively small in the earlier time periods of our data set. If a fixed-size volume bar is applied, each of the initial volume bars contains five to ten days of data. Thus, for certain countries the BV–VPIN values are unavailable for the initial months or years in our data set.

We find that three countries have BV–VPIN values associated with the AFC, and only in Japan does BV–VPIN exhibit some safe haven properties. During the DCB, all countries except Australia exhibit safe haven properties. During the GR, BV–VPIN exhibits safe haven qualities in four countries, particularly in the United States and Germany, where the coefficients are statistically significant at 1%. In the United Kingdom and China, BV–VPIN is not an effective safe haven during the GR. BV–VPIN

TABLE 6 Hedge and safe haven properties of BV–VPIN during periods of financial stress.

	Hedge		1997 AFC		2000 DCB		2008 GR		2011 USCRD	
Country	**Coeff.**	***t*-stats**	**Ttl. eff.**	***t*-stats**	**Ttl. eff.**	***t*-stats**	**Ttl. eff.**	***t*-stats**	**Ttl. eff.**	***t*-stats**
United States	−0.004	−0.365	0.022	0.123	−0.055	−0.202	−0.150***	−2.798	-0.192***	−2.861
United Kingdom	0.004	0.279	0.353	2.014	−0.170	−0.570	0.503	16.880	−0.302***	−2.841
Germany	−0.010	−1.173	N/A	N/A	−0.020	−0.106	−0.408***	−6.907	-0.100**	−2.133
Japan	−0.019**	−2.139	−0.087	−1.008	−0.053	−0.198	−0.056	−0.766	0.221	2.620
China	0.003	0.227	N/A	N/A	N/A	N/A	0.088	0.782	−0.224**	−2.493
Australia	−0.035**	−2.118	N/A	N/A	0.236	0.759	−0.083	−0.983	−0.140	−1.147

The estimation results for the coefficients in (4.7). The duration of crisis periods is set at twenty days after the crisis starts. Negative coefficients in the hedge column (c_0) signify that BV–VPIN is a hedge against the market. Negative coefficients in subsequent columns show that it is a safe haven during the AFC (c_1), DCB (c_2) GR (c_3) or USCRD (c_4). Each t-statistics column is associated with the coefficient column to the left as an indication of the significance level. *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

is a strong safe haven in all countries except for Japan during the period around the USCRD. It exhibits statistical significance in the United States, the United Kingdom, Germany and Australia. In four countries out of six, BV–VPIN exhibits hedging qualities. In both Japan and Australia, significant negative coefficients are observed, thus indicating strong hedge properties against market crises. These results show that if the BV–VPIN were to be operationalized as a futures contract, it would exhibit suitable safe haven and hedge characteristics for asset managers, who wish to protect themselves from the high volatility and liquidity shocks that may occur due to order imbalances.

4.4 BV–VPIN in a risk-on/risk-off trading strategy

We test a risk-on/risk-off trading strategy based on the CDF of BV–VPIN. Investments are in equities (risk-on) when the CDF of BV–VPIN is below a set threshold; otherwise, we invest in the risk-free rate (risk-off). Equities are represented by each country's stock index, and the risk-free rate used is the US ten-year treasury rate. We refer to this approach as the BV–VPIN strategy. For comparable benchmark strategies, we report the realized returns on a buy-and-hold investment in (1) equities and (2) risk-free rate.

Table 7 represents the performance of these three strategies for each of our six different countries. Similar to the analysis of prediction accuracy in Section 4.2, threshold values of 0.6, 0.8 and 0.9 are applied. We report descriptive statistics such as the mean, standard deviation, skewness and kurtosis. Risk-adjusted returns performance is given by the Sharpe ratio (Sharpe 1994) and the Sortino ratio (Sortino and Satchell 2001), which capture total volatility and downside volatility, respectively. The initial investment for each approach is assumed to be US$100, and we report the final value (FV) and total return (TR) in dollars and as percentages, respectively.

Across all countries – with the exception of Germany – every strategy exhibits negative skewness, with the highest being the BV–VPIN strategies at a threshold of 0.6 and the lowest being the risk-free strategy. Across all countries, the BV–VPIN and equities strategies exhibit excess kurtosis; the risk-free strategy does not. Changes to sentence OK?

For all six countries, investment in the risk-free rate produces the highest Sharpe ratio due to its low values of standard deviation. The BV–VPIN strategy with an optimal threshold outperforms the equities strategy. In the US market, the BV–VPIN strategy with a threshold of 0.9 has the highest Sharpe ratio. In China, the Sharpe ratio using the BV–VPIN strategy is double that of the equities strategy. In Japan, the BV–VPIN strategy with a threshold of 0.6 produces a Sharpe ratio five times larger than that of the equities strategy. As measured by the Sortino ratio, the BV–VPIN strategy with an optimal threshold also outperforms equities in all countries with the exception of the United States. In the US market, the BV–VPIN strategy has the same

Changes to sentence OK?

Words added – OK?

TABLE 7 Evaluation of BV–VPIN in trading strategy. [Table continues on next page.]

(a) United States

Strategy	Thres.	Mean	SD	Skew	Kurt.	Sharpe	Sortino	FV ($)	TR (%)
BV–VPIN	0.9	0.026	1.123	−0.158	7.363	0.024	0.033	263	163.17
	0.8	0.020	0.941	−0.205	5.907	0.021	0.029	207	107.41
	0.6	0.015	0.661	−0.435	10.382	0.023	0.032	187	86.51
Equities		0.029	1.252	−0.043	7.616	0.023	0.033	273	172.97
Risk-free		0.011	0.004	−0.025	−0.934	2.974	N/A	169	69.11

(b) United Kingdom

Strategy	Thres.	Mean	SD	Skew	Kurt.	Sharpe	Sortino	FV ($)	TR (%)
BV–VPIN	0.9	0.012	1.099	−0.132	6.752	0.011	0.016	135	34.82
	0.8	0.007	0.992	−0.343	7.571	0.007	0.010	112	11.71
	0.6	0.013	0.619	−0.314	13.662	0.022	0.030	172	71.87
Equities		0.014	1.224	−0.009	5.491	0.011	0.016	134	34.02
Risk-free		0.011	0.004	−0.069	−0.979	3.030	N/A	165	65.34

(c) Germany

Strategy	Thres.	Mean	SD	Skew	Kurt.	Sharpe	Sortino	FV ($)	TR (%)
BV–VPIN	0.9	0.028	1.407	0.152	6.867	0.020	0.029	218	118.50
	0.8	0.034	1.195	0.098	9.777	0.028	0.040	310	209.68
	0.6	0.021	0.869	−0.281	15.859	0.024	0.034	210	109.68
Equities		0.030	1.535	0.115	4.432	0.019	0.027	214	113.59
Risk-free		0.010	0.003	0.036	−0.874	3.024	N/A	155	55.22

Sortino ratio as the equities strategy. Generally, we find that the BV–VPIN strategy results in higher risk-adjusted returns.

When analyzing the FV and TR of each strategy, application of the optimal BV–VPIN threshold outperforms equities and the risk-free rate in all countries with the exception of the United States. With a threshold of 0.9, the FV of the BV–VPIN strategy comes close to that of the equities strategy for the United States. For the United Kingdom, equities generally performed poorly compared with investing in the risk-free rate. However, with a threshold of 0.6, the BV–VPIN strategy outperforms both equities and the risk-free rate. In Germany, the BV–VPIN strategy at thresholds 0.9 and 0.8 successfully outperforms both equities and the risk-free strategies. In the Japanese market, equities performs poorly compared with the risk-free rate. However, applying the BV–VPIN strategy results in 307% returns with a threshold of 0.6, which

Words added – OK?

TABLE 7 Continued.

(d) Japan

Strategy	Thres.	Mean	SD	Skew	Kurt.	Sharpe	Sortino	FV ($)	TR (%)
BV–VPIN	0.9	0.029	1.243	−0.409	11.415	0.023	0.032	262	162.24
	0.8	0.021	1.100	−0.596	6.597	0.019	0.026	193	92.55
	0.6	0.035	0.891	−0.750	11.036	0.039	0.055	407	307.24
Equities		0.011	1.553	−0.146	5.440	0.007	0.010	96	−3.65
Risk-free		0.011	0.004	−0.085	−0.976	3.044	N/A	163	62.50

(e) China

Strategy	Thres.	Mean	SD	Skew	Kurt.	Sharpe	Sortino	FV ($)	TR (%)
BV–VPIN	0.9	0.075	1.379	−0.364	4.091	0.054	0.078	773	673.27
	0.8	0.062	1.218	−0.224	5.211	0.051	0.074	562	461.98
	0.6	0.053	0.930	−0.239	7.927	0.057	0.084	463	363.61
Equities		0.044	1.686	−0.287	3.716	0.026	0.036	255	154.90
Risk-free		0.009	0.003	−0.100	−1.258	3.306	N/A	133	33.14

(f) Australia

Strategy	Thres.	Mean	SD	Skew	Kurt.	Sharpe	Sortino	FV ($)	TR (%)
BV–VPIN	0.9	0.031	0.829	−0.251	3.003	0.037	0.052	346	246.05
	0.8	0.031	0.772	−0.346	3.168	0.040	0.057	361	260.79
	0.6	0.017	0.577	−0.583	6.035	0.029	0.040	198	98.17
Equities		0.021	0.967	−0.431	5.317	0.022	0.030	209	108.99
Risk-free		0.011	0.003	−0.081	−0.958	3.069	N/A	162	62.08

A comparison of the performance of investing in (1) BV–VPIN, a risk-on/risk-off strategy that invests in the respective country's equity index (risk-free rate) when the BV–VPIN is below (above) a set threshold; (2) equities, a buy-and-hold strategy in each country's stock index; and (3) risk-free, a buy-and-hold strategy in the form of a US ten-year treasury rate. The distribution characteristics of each (eg, mean, standard deviation (SD), skewness and kurtosis) for each strategy are reported. The risk-adjusted returns are reported with the Sharpe and Sortino ratios. The final value of investment by investing an initial sum of US$100 is given by FV ($), and TR (%) is the total return of the strategy. The Sortino ratio for the risk-free rate is not applicable, as the risk-free rate is never negative in our data set.

is five times greater than the return from investing in the risk-free rate. The BV–VPIN strategy, regardless of threshold value, successfully outperforms equities and the risk-free rate in China. Using a threshold value of 0.9, the BV–VPIN strategy generates a fourfold return relative to equities; this becomes a threefold return for threshold 0.8 and a twofold return for threshold 0.6. In Australia, application of the BV–VPIN strategy doubles the return on investment compared with the equities strategy. Change OK?

Figure 5 shows the accumulation of wealth with an initial investment of US$100 dollars using the equities and BV–VPIN strategies with different thresholds. We find that the BV–VPIN strategy effectively avoids large losses during extreme market downturns. For China and Australia, the BV–VPIN strategy suggests investing in the risk-free asset before the market drops and reinvesting in equities as the market starts to recover. Specifically, we observe that in China, Australia, the United Kingdom and Japan the BV–VPIN avoids investing in equities during the GR period.

The optimal threshold to be applied for BV–VPIN varies across different markets. For China and the United States, 0.9 is optimal, while in Australia and Germany 0.8 produces the highest returns. In the United Kingdom and Japan, 0.6 significantly outperforms higher thresholds. Our results indicate that one should tailor the choice of optimal threshold depending on the market. However, the application of a threshold value of 0.8 seems to result in a robust and profitable performance across all markets investigated in our study.

5 CONCLUSION

The occurrence of informed trading may have severe negative impacts for market makers and stock exchanges (Sun and Ibikunle 2017; Wong *et al* 2009). A structural model for market microstructure research proposed by Easley *et al* (1996) provides us with the estimation of PIN. The importance of PIN is that it provides a measure of order flow toxicity, which can have a negative impact on market liquidity. With the advent of HFT, Easley *et al* (2012) develop a modified measure, BV–VPIN, where the probability of informed trading is based on volume imbalance and trade intensity.

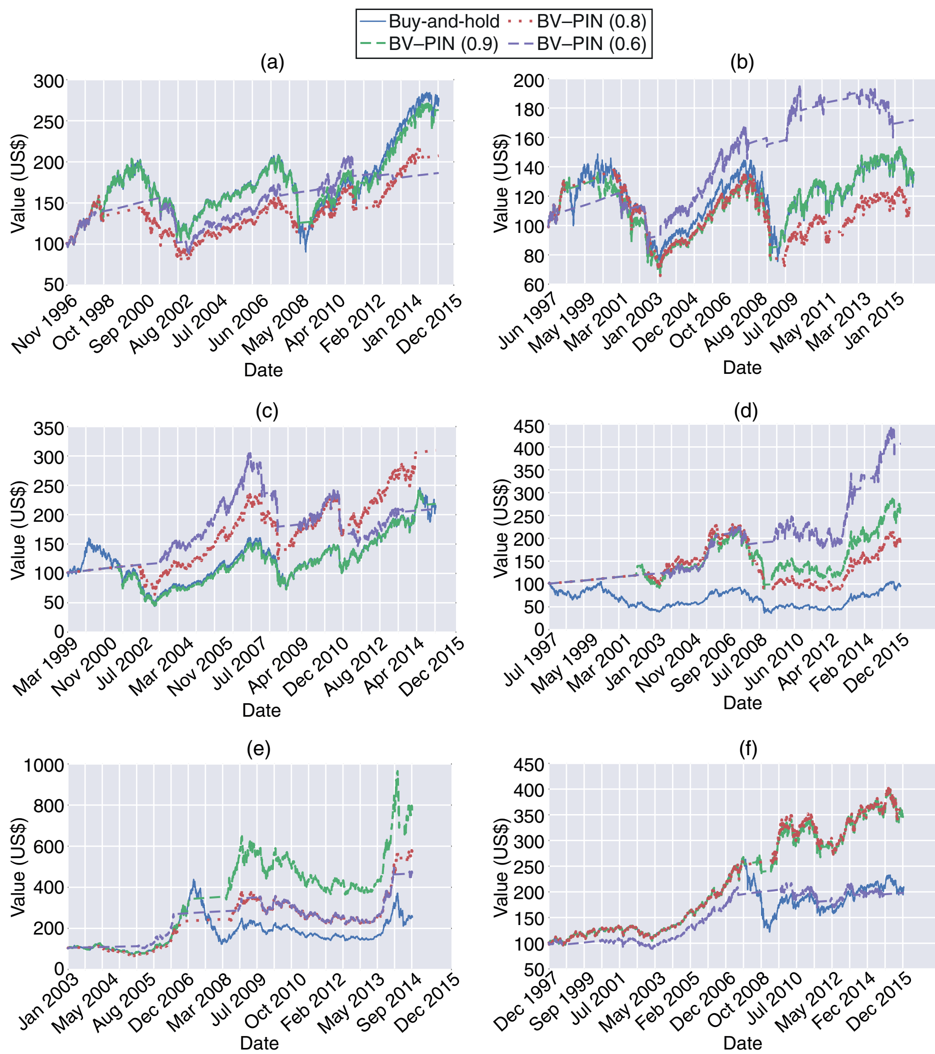
Change OK?

Market makers' estimates of time-varying toxicity levels are a crucial factor in determining their participation, such that, if they believe that toxicity is high, they will liquidate their positions and leave the market, thus affecting market liquidity. Henceforth, BV–VPIN is a procedure that estimates the probability of informed trading based on volume imbalance and trade intensity.

Remove this sentence? It's pretty much identical to the one above!

Empirical tests suggest that BV–VPIN functions not only as an indicator of current market liquidity but also as an early warning signal for impending volatility. Although BV–VPIN itself is not effective in predicting future price movements, the CDF of BV–VPIN exhibits predictive properties in an HFT environment (Easley *et al* 2011b). Abad and Yagüe (2012) propose that alternative specifications of the BV–VPIN model can result in alternative gauges for adverse selection. Thus, they suggest that BV–VPIN can be a broad measure of adverse selection beyond that of HFT applications. We apply BV–VPIN on daily equity data across a range of international equity indexes. We select optimal parameters for BV–VPIN based on the US market and apply it across a range of major global economies to ascertain its capabilities as a long-term indicator of market liquidity and impending market downturns. Our work includes the analysis of BV–VPIN's role as a safe haven and hedge instrument if implemented as a futures contract, as suggested by Easley *et al* (2011a). To investigate the economic

FIGURE 5 Accumulation of wealth using different trading strategies for international equity indexes.



(a) United States. (b) United Kingdom. (c) Germany. (d) Japan. (e) China. (f) Australia.

significance of BV–VPIN as an early warning signal, we apply it in a risk-on/risk-off trading strategy, where an asset manager invests in equities (risk-free rate) when the CDF of BV–VPIN is below (above) a threshold value.

We find the CDF of BV–VPIN exhibits predictive properties; however, the choice of threshold value can vary across different countries. This is in line with the findings of Easley *et al* (2014), who contend that the selection of parameters of BV–VPIN needs to cater for the characteristics of the instrument being investigated. Our investigation of the hedge and safe haven properties of BV–VPIN suggests that it is an effective safe haven for most countries during extreme market downturns. BV–VPIN exhibits strong safe haven characteristics for most markets during three major crisis periods: the 2000 DCB, the 2008 GR and the 2011 USCRD. Incorporating BV–VPIN as part of a flight-to-quality trading strategy results in an outperformance of the buy-and-hold equities strategy for four out of the six countries in our study. Specifically, in China, Japan and Australia, the BV–VPIN strategy doubles the returns of a comparable buy-and-hold strategy in equities. In conclusion, we show that BV–VPIN has applications beyond that of HFT environments. BV–VPIN is an effective indicator of long-term market volatility and exhibits hedging and safe haven characteristics during market downturns. BV–VPIN may also be used to improve portfolio management strategies for asset managers, ascertaining market conditions in which one should minimize risk by reducing exposure to equities. For future work, other applications of BV–VPIN could involve investigating its ability to track the "smart money" effect, as documented in the managed mutual funds industry (Gruber 1996; Sapp and Tiwari 2004; Zheng 1999), and tracking money flows between asset classes such as the flight-to-quality effect (equity to debt) by mutual fund managers.

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

ACKNOWLEDGEMENTS

The authors acknowledge the resources and technical support provided by the research team at PF2 Securities. We also thank Joe Pimbley at Maxwell Consulting for his insightful feedback and helpful comments. The authors acknowledge funding from (1) the Australia Awards – Endeavour Research Fellowship and (2) the Australian Institute for Business and Economics (AIBE) in support of Dr. Rand K. Y. Low at the Stern School of Business, New York University. The authors also acknowledge financial support for this research from the following funding bodies: (1) Accounting & Finance Association of Australia and New Zealand (AFAANZ) 2013/2014 research grants, and (2) a UQ Postdoctoral Research Grant titled "Portfolio optimization and risk management techniques for financial crisis".

REFERENCES

Abad, D., and Yagüe, J. (2012). From PIN to VPIN: an introduction to order flow toxicity. *Spanish Review of Financial Economics* **10**(2), 74–83.

Andersen, T. G., and Bondarenko, O. (2014a). Reflecting on the VPIN dispute. *Journal of Financial Markets* **17**, 53–64.

Andersen, T. G., and Bondarenko, O. (2014b). VPIN and the flash crash. *Journal of Financial Markets* **17**, 1–46.

Andersen, T. G., and Bondarenko, O. (2015). Assessing measures of order flow toxicity and early warning signals for market turbulence. *Review of Finance* **19**(1), 1–54.

Ang, A., and Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of Financial Economics* **63**(3), 443–494.

Aslan, H., Easley, D., Hvidkjaer, S., and O'Hara, M. (2011). The characteristics of informed trading: implications for asset pricing. *Journal of Empirical Finance* **18**(5), 782–801.

Baker, H. K., and Kiymaz, H. (2013). *Market Microstructure in Emerging and Developed Markets*. Wiley.

Baur, D. G., and McDermott, T. K. (2010). Is gold a safe haven? International evidence. *Journal of Banking & Finance* **34**(8), 1886–1898.

Bell, H. A. (2013). High frequency trading: do regulators need to control this tool of informationally efficient markets? Policy Analysis 731, Cato Institute.

Chakrabarty, B., Pascual, R., and Shkilko, A. (2012). Trade classification algorithms: a horse race between the bulk-based and the tick-based rules. Working Paper 2182819, Social Science Research Network.

Chen, Y., and Zhao, H. (2012). Informed trading, information uncertainty, and price momentum. Journal of Banking & Finance 36 (7), 2095–2109.

Corcoran, C. M. (2012). *Systemic Liquidity Risk and Bipolar Markets: Wealth Management in Today's Macro Risk On/Risk Off Financial Environment*. Wiley.

Easley, D., Kiefer, N. M., O'Hara, M., and Paperman, J. B. (1996). Liquidity, information, and infrequently traded stocks. *Journal of Finance* **51**(4), 1405–1436.

Easley, D., Kiefer, N. M., and O'Hara, M. (1997). The information content of the trading process. *Journal of Empirical Finance* **4**(2), 159–186.

Easley, D., Engle, R. F., O'Hara, M., and Wu, L. (2008). Time-varying arrival rates of informed and uninformed trades. *Journal of Financial Econometrics* **6**(2), 171–207.

Easley, D., Hvidkjaer, S., and O'Hara, M. (2010). Factoring information into returns. *Journal of Financial and Quantitative Analysis* **45**, 293–309.

Easley, D., de Prado, M. L., and O'Hara, M. (2011a). The exchange of flow toxicity. *Journal of Trading* **6**(2), 8–13.

Easley, D., de Prado, M. M. L., and O'Hara, M. (2011b). The microstructure of the "flash crash": flow toxicity, liquidity crashes, and the probability of informed trading. *Journal of Portfolio Management* **37**(2), 118–? Please provide missing page number.

Easley, D., de Prado, M. M. L., and O'Hara, M. (2012). Flow toxicity and liquidity in a high-frequency world. *Review of Financial Studies* **25**(5), 1457–1493.

Easley, D., de Prado, M. M. L., and O'Hara, M. (2014). VPIN and the flash crash: a rejoinder. *Journal of Financial Markets* **17**, 47–52.

Easley, D., Prado, M. L., and O'Hara, M. (2015). Optimal execution horizon. *Mathematical Finance* **25**(3), 640–672.

Easley, D., de Prado, M. L., and O'Hara, M. (2016). Discerning information from trade data. *Journal of Financial Economics* **120**(2), 269–285.

Flood, J. (2010). NYSE confirms price reporting delays that contributed to the flash crash. *Chief Investment Officer*, August 24.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48**(5), 1779–1801.

Gruber, M. J. (1996). Another puzzle: the growth in actively managed mutual funds. *Journal Of Finance* **51**(3), 783–810.

Kang, M. (2010). Probability of information-based trading and the January effect. *Journal of Banking & Finance* **34**(12), 2985–2994.

Krasting, B. (2010). Then yen did it? *Seeking Alpha*, May 7.

Lin, H.-W. W., and Ke, W.-C. (2011). A computing bias in estimating the probability of informed trading. *Journal of Financial Markets* **14**(4), 625–640.

Longin, F., and Solnik, B. (1995). Is the correlation in international equity returns constant: 1960–1990? *Journal of International Money and Finance* **14**(1), 3–26.

Low, R. K. Y., Alcock, J., Faff, R., and Brailsford, T. (2013). Canonical vine copulas in the context of modern portfolio management: are they worth it? *Journal of Banking & Finance* **37**(8), 3085–3099.

Low, R. K. Y., Yao, Y., and Faff, R. (2016). Diamonds vs. precious metals: what shines brightest in your investment portfolio? *International Review of Financial Analysis* **43**, 1–14.

Patterson, S., and Lauricella, T. (2010). Did a big bet help trigger "Black Swan" stock swoon? *Wall Street Journal*, May 10.

Phillips, M. (2010). SEC's Schapiro: here's my timeline of the flash crash. *Wall Street Journal*, May 20.

Sapp, T., and Tiwari, A. (2004). Does stock return momentum explain the "smart money" effect? *Journal of Finance* **59**(6), 2605–2622.

Securities and Exchange Commission and Commodity Futures Trading Commission (2010a). Findings regarding the market events of May 6, 2010. Report, September, CFTC–SEC.

Securities and Exchange Commission and Commodity Futures Trading Commission (2010b). Preliminary findings regarding the market events of May 6, 2010. Report, May, CFTC–SEC.

Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management* **21**(1), 49–59.

Sortino, F. A., and Satchell, S. (2001). *Managing Downside Risk in Financial Markets: Theory, Practice and Implementation*. Butterworth-Heinemann.

Sun, Y., and Ibikunle, G. (2017). Informed trading and the price impact of block trades: a high frequency trading analysis. *International Review of Financial Analysis* **54**, 114–129.

Vega, C. (2006). Stock price reaction to public and private information. *Journal of Financial Economics* **82**(1), 103–133.

Wong, W. K., Tan, D., and Tian, Y. (2009). Informed trading and liquidity in the Shanghai stock exchange. *International Review of Financial Analysis* **18**(1–2), 66–73.

Wu, K., Bethel, E., Gu, M., Leinweber, D., and Rübel, O. (2013). A big data approach to analyzing market volatility. *Algorithmic Finance* **2**(3–4), 241–267.

Zheng, L. (1999). Is money smart? A study of mutual fund investors' fund selection ability. *Journal of Finance* **54**(3), 901–933.

Zweig, J. (2012). Could computers protect the market from computers? *Wall Street Journal*, May ???.

Please provide date.