Semiparametric Estimation of the Bid-Ask Spread in Extended Roll Models

Xiaohong Chen[a*] Oliver Linton[b†] Stefan Schneeberger[c‡] Yanping Yi[d §]

[a] Cowles Foundation for Research in Economics, Yale University

[b] Department of Economics, University of Cambridge

[c] Department of Economics, Yale University

[d] School of Economics, Shanghai University of Finance and Economics

First version: August 2015; Revised November 22, 2017

Abstract

We propose new methods for estimating the bid-ask spread from observed transaction prices alone. Our methods are based on the empirical characteristic function. We compare our methods theoretically and numerically with the Roll (1984) method as well as with its best known competitor, the Hasbrouck (2004) method, and find that our estimators perform much better when this distribution is far from Gaussian. Our methods are applied to the E-mini futures contract on the S&P 500 during the Flash Crash of May 6, 2010. We also establish $\sqrt{T}$ consistency and asymptotic normality of the proposed estimators in various extended Roll models.

JEL Classification Number: C12, C13, C14.

Keywords: Bid-ask spread; Roll model; Semiparametric estimation; Empirical characteristic function; Latent variables.

*PO Box 208281, New Haven CT 06520-8281, USA. E-mail: xiaohong.chen@yale.edu.

†Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, UK. E-mail: obl20@cam.ac.uk.

‡New Haven CT 06520-8281, USA. E-mail: stefan.schneeberger@yale.edu.

§The corresponding author. Tel.: (+86)21-65902962. Address : School of Economics, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai, 200433, China. E-mail: yi.yanping@mail.shufe.edu.cn.

1 Introduction

The bid-ask spread of a financial asset is the difference between the ask and the bid quotes. The spread reflects the cost of providing market liquidity, the difference in price paid by an urgent buyer and received by an urgent seller, which is a major part of the transaction cost facing investors. It has been studied extensively by financial economists, see, e.g. Glosten and Milgrom (1985), Glosten and Harris (1988), Harris (1990), Huang and Stoll (1997), Schultz (2001), Harris and Piwowar (2006), Corwin and Schultz (2012), Bleaney and Li (2015), and the references therein. The estimation strategy of the bid-ask spread (transaction cost) depends crucially on the market structure and *the data availability*.

Measuring the bid-ask spread in practice can be quite time consuming (reconstruction of the limit order book is required) and may be subject to a number of potential accuracy issues due to the quoting strategies of High Frequency Traders, for example. For the U.S. municipal bond market (see, e.g., Harris and Piwowar (2006)) and the U.S. corporate bond market (see, e.g., Edwards et al. (2007)), the firm bid and ask quotes are absent. As for other over-the-counter (OTC) markets, market-wide transaction data are generally not available (see, e.g., Jankowitsch et al. (2011)). Data are also limited for open outcry markets (e.g. the futures trading in CME), where bid and ask quotes by traders expire (if not filled) without recording (see, e.g., Hasbrouck (2004)). Moreover, in the U.S. markets transaction data are only available since 1983 and in many countries transaction data are not available at all.

Using observed transaction prices alone, the seminal paper Roll (1984) proposes a simple model to estimate the effective bid-ask spread without information on the bid and the ask quotes, or the trade direction (i.e., whether the trade initiator is a buyer or a seller). The basic Roll estimator has seen its popularity in analyzing the U.S. historical data sets (prior to 1983), the international markets without transaction data, the illiquid markets (particularly OTC markets), the cases when intraday quotes and trades cannot be reliably matched, and the cases when the transaction data are cumbersome to use or expensive to purchase.

In the Roll (1984) model, an observed (log) asset price p_t evolves according to

$$p_t = p_t^* + I_t \frac{s_0}{2}, \quad p_t^* = p_{t-1}^* + \varepsilon_t, \tag{1}$$

$$\Delta p_t = \varepsilon_t + (I_t - I_{t-1}) \frac{s_0}{2} = \varepsilon_t + \Delta I_t \frac{s_0}{2}, \tag{2}$$

where $\{p_t^*\}$ are the underlying fundamental (log) price with serially uncorrelated innovations $\{\varepsilon_t\}$. The trade direction indicators $\{I_t\}$ are i.i.d. and take the values $+1$ (if the transaction is buyer initiated), or -1 (if the transaction is seller initiated) with equal probability. $\{\varepsilon_t\}$ are uncorrelated with $\{I_t\}$. Essentially, Roll (1984) assumes an informationally efficient market. The parameter of interest s_0 is the effective bid-ask spread, measuring the order processing cost. *The transaction prices* $\{p_t\}$ *are the only observable variables in Eq.*(1). Thus, assuming the one-period returns $\{\Delta p_t\}$ have finite second moments, the true unknown s_0 is identified using the population auto-covariance of $\{\Delta p_t\}$ and can be estimated using its sample analogue

$$s_0 = 2\sqrt{-\operatorname{Cov}(\Delta p_t, \Delta p_{t-1})}, \quad \widehat{s}_{Roll} := 2\sqrt{-\widehat{\operatorname{Cov}}(\Delta p_t, \Delta p_{t-1})}. \tag{3}$$

In practice, this estimator is not satisfactory, since the empirical first-order autocovariance of one-period returns is often positive, then Eq.(3) is not well-defined. Roll (1984) encounters this phenomenon in about a half of the cases in his data, which consists of annual samples of daily and weekly prices. The literature contains several proposals to deal with this shortcoming. Harris (1990) suggests to replace $-\widehat{\operatorname{Cov}}(\Delta p_t, \Delta p_{t-1})$ in (3) by its absolute value $\left|\widehat{\operatorname{Cov}}(\Delta p_t, \Delta p_{t-1})\right|$. This makes the estimator always well-defined. Hasbrouck (2009) suggests to set the estimated spread to zero if the empirical autocovariance is positive, which is motivated by the finding of Harris (1990) that positive autocovariance estimates are more likely for smaller spreads. However, it is not clear whether either of these ad hoc modifications work well in finite samples, and they are theoretically not well motivated.

A well-known alternative by Hasbrouck (2004) proposes to estimate the bid-ask spread based on Bayesian analysis, using the Gibbs sampler. In doing so, he uses a stronger version of the Roll model, in which $\varepsilon_t \sim$ i.i.d. $\mathrm{N}\left(0, \sigma_\varepsilon^2\right)$ and is independent of $\{I_t\}$. The unknown parameter σ_ε is estimated jointly with the spread s_0. Unfortunately the Hasbrouck (2004) estimator performs poorly or even is not well defined when the distribution of ε_t is far from Gaussian, e.g. fat-tailed or asymmetric. However, the Gaussian assumption generally fails in financial data. Corwin and Schultz (2012) develop another spread estimator from consecutive daily high and low transaction prices. They also assume that the fundamental price process is a geometric Brownian motion, which is even stronger than the discrete time Gaussian assumption employed in Hasbrouck (2004).

The recent empirical literature emphasizes several issues with the Roll model : (a) Market orders are assumed not to bring news into prices, so that I_t has no effect on the underlying true price p_t^*. However, the literature finds the quoted prices increase after a buyer-initiated trade (see, e.g., Glosten and Milgrom (1985), Glosten and Harris (1988), Huang and Stoll (1997), Muravyev (2016)). (b) It assumes balanced market order flow, i.e., $q = 1/2$, which may be accurate on average, but may be inaccurate for certain episodes of trading (see, e.g., Brunnermeier and Pedersen (2005), Ito and Yamada (2016)). (c) It assumes always a price change due to transactions, but many transactions might happen with no price change (see, e.g., Huang and Stoll (1997)). In the presence of any of these effects, one is not able to identify the spread jointly with parameters describing adverse selection cost or order flow imbalance, using either Roll (1984)'s or Hasbrouck (2004)'s methodology, without additional assumptions or observed information. The spread estimators of Roll (1984) and Hasbrouck (2004) will also be inconsistent. There have been many recent suggestions for estimating spreads (and liquidity costs more generally), that relax some of these assumptions, but at the cost of requiring additional observed information (data) such as trade direction indicators. As we have mentioned, these data may not be readily available or, if available, be not well measured for the relevant frequency (see, e.g., Andersen and Bondarenko (2014)). Bleaney and Li (2015) provide a detailed discussion of all the above and additional problems with the basic Roll (1984) model. Goyenko et al. (2009) review many different liquidity proxies based on lower frequency data, including the Roll-type transaction-price-based measures, as well as those that use additional information such as trading volumes.

In this paper, we work with the framework in Eq.(1) and its simple extensions, where *only* *transaction prices are available.* These prices could be daily or weekly closing prices, but might also consist of high frequency intraday prices. However, contrary to, e.g., Corwin and Schultz (2012), we do not require intraday data for our method to work. We assume that $\{\varepsilon_t\}$ is i.i.d. and independent of the increments of the unobserved trade direction indicators $\{\Delta I_t\}$, or independent of $\{I_t\}$ when adverse selection cost is considered. The independence assumption allows us to propose new, simple estimators of s_0 that are based on empirical characteristic functions. However, we do not impose any parametric restrictions (in contrast to Hasbrouck (2004)), or any location/scale assumptions, and we do not require the existence of moments of any order (in contrast to Roll (1984), which requires ε_t to have finite second moments). This feature seems to be attractive for financial applications where distributions can be asymmetric and heavy-tailed. In addition to the basic Roll (1984) model, we also propose solutions to the three problems (a)-(c) with the Roll model listed above. We show how to estimate parameters that capture an adverse selection component in the spread in Section 3, or those associated with unbalanced order flow in Section 4, or those that characterize the probability of no price change in Section 5. The consistency and asymptotic normality of our estimators are established without requiring finite moments of the observed price data. In simulation studies that mimic the design of Hasbrouck (2009), our estimators are competitive to Roll (1984)'s and Hasbrouck (2004)'s when the latent true fundamental return distribution is Gaussian, and perform much better when the distribution is either asymmetric or heavy-tailed.

We apply our estimators to a high-frequency dataset of transaction prices on the E-mini futures contract during the Flash Crash of May 6, 2010. We use a rolling-window approach to understand the development of the spread during the crisis period and more tranquil periods. In the application, we also show the evolution of some additional estimated quantities, including the estimated characteristic function of the fundamental price innovations ε_t, indicators for an unbalanced order flow, an adverse selection component in the spread, and the aggregation robustness of our method.

The rest of the paper is organized as follows: Section 2 presents the basic model and provides new simple spread estimators and their asymptotic properties. In Section 3, we study the estimation of adverse selection cost. In Section 4, we address order flow imbalance in a simple extended model. In Section 5, a simplified model is used to consider the case when the transaction may occur without price change. Section 6 presents a simulation study and the empirical application. Section 7 concludes. All the proofs, some figures and tables are presented in the online supplement.

2 Basic Model and Large Sample Properties of Estimators

In this section we assume that the observed price dynamics follow a basic Roll (1984) type model.

Assumption 1. *(i) Data $\{p_t\}_{t=1}^T$ is generated from Eq.* (1) *with $s_0 > 0$, where $\{\varepsilon_t\}$ is i.i.d. and independent of $\{\Delta I_t\}$ and has unknown distribution function F_ϵ; (ii) $\{I_t\}$ is i.i.d.; and (iii) I_t takes the values ± 1 with equal probability.*

The distribution of ε_t could be continuous or discrete and could have no finite moments. Let $\varphi_\varepsilon(u) := \mathbb{E}(\exp(iu\varepsilon_t))$ denote the characteristic function (c.f.) of ε_t. Let $\varphi_{\Delta p,2}(u,u') := \mathbb{E}(\exp(iu\Delta p_t + iu'\Delta p_{t-1}))$ and $\varphi_{\Delta p,1}(u) := \mathbb{E}(\exp(iu\Delta p_t)) = \varphi_{\Delta p,2}(u,0)$ denote the joint c.f. of $(\Delta p_t, \Delta p_{t-1})$ and the marginal c.f. of Δp_t, respectively. By definition, they are nonparametrically identified and estimable from data. We shall obtain a useful expression based on these quantities that will identify the unknown spread parameter $s_0 > 0$. The use of marginal quantities such as characteristic functions for identification of s_0 is reminiscent of the classic GMM approach to identification and estimation of continuous time models where the transition density is hard to express analytically, but many moment conditions can be obtained from the marginal distributions. Precisely, Assumption 1 implies that, for all $(u,u') \in \mathbb{R}^2$,

$$\varphi_{\Delta p,2}(u,u') = \varphi_\varepsilon(u)\varphi_\varepsilon(u')\cos\left(u\frac{s_0}{2}\right)\cos\left((u'-u)\frac{s_0}{2}\right)\cos\left(u'\frac{s_0}{2}\right). \tag{4}$$

If the distribution of ε_t were parametrically specified, one could work directly with equation (4) and develop estimation methods that would be a simple alternative to the Hasbrouck (2004) likelihood-type procedure. In our case, where this distribution is not specified, these relations still involve the unknown function φ_ε, albeit in a convenient multiplicative fashion. We find a relation that eliminates the unknown function $\varphi_\varepsilon(\cdot)$, and then proceed to estimate the parametric model for the trade direction effect s_0. Denote

$$\overline{\mathcal{V}} := \{u \in \mathbb{R} : \varphi_{\Delta p,1}(u) \neq 0\}. \tag{5}$$

Since $\varphi_{\Delta p,1}(\cdot)$ is uniformly continuous in $\mathbb{R}$ (see, e.g., page 3 of Lukacs (1972)) and $\varphi_{\Delta p,1}(0) = 1$, $\overline{\mathcal{V}}$ contains an open interval of 0[1]. Denote

$$H(u,u') := \frac{\varphi_{\Delta p,2}(u,u')}{\varphi_{\Delta p,1}(u)\varphi_{\Delta p,1}(u')} \quad \textit{for any } (u,u') \in \overline{\mathcal{V}}^2, \tag{6}$$

which is nonparametrically estimable from the data $\{\Delta p_t\}$. Eq. (4) implies that

$$H(u,u') = \frac{\cos\left((u-u')\frac{s_0}{2}\right)}{\cos\left(u\frac{s_0}{2}\right)\cos\left(u'\frac{s_0}{2}\right)} =: R(u,u';s_0) \quad \textit{for all } (u,u') \in \overline{\mathcal{V}}^2, \tag{7}$$

and therefore $H(u,u')$ is real-valued for all $(u,u') \in \overline{\mathcal{V}}^2$. Or equivalently,

$$\varphi_{\Delta p,2}(u,u') = \varphi_{\Delta p,1}(u)\varphi_{\Delta p,1}(u')R(u,u';s_0) \quad \textit{for all } (u,u') \in \overline{\mathcal{V}}^2. \tag{8}$$

Eq. (7) (or (8)) is free of the nuisance function $\varphi_\varepsilon(\cdot)$ and only depends on the parameter of interest s_0. Chen et al. (2017) obtains the identification result for s_0 and the c.f. $\varphi_\varepsilon(\cdot)$ using either the diagonal information or the off-diagonal information of Eq. (7) (or (8)).

Eq. (7) (or (8)) for estimation of s_0 is similar to the classic GMM approach to estimation. Due to the continuity of the c.f. $\varphi_{\Delta p,2}(u,u')$ in $\mathbb{R}^2$ and $\varphi_{\Delta p,2}(0,0) = 1$, $\overline{\mathcal{V}}^2$ contains an open ball of $(0,0)$, and hence Eq. (7) (or (8)) contains infinitely many overidentifying restrictions for s_0. Let $\mathcal{S} := [0, \bar{s}]$ denote the parameter space, where $\bar{s} > 0$ is chosen from prior experience for the market (to ensure that $s_0 \in \mathcal{S}$). Denote

$$\overline{\mathcal{U}} := \left\{(u,u') \in \overline{\mathcal{V}}^2 : \min_{s\in\mathcal{S}} \left|\cos\left(u\frac{s}{2}\right)\cos\left(u'\frac{s}{2}\right)\right| > 0\right\}, \tag{9}$$

which still contains an open ball of $(0,0)$. Denote

$$R(u,u';s) := \frac{\cos\left((u-u')\frac{s}{2}\right)}{\cos\left(u\frac{s}{2}\right)\cos\left(u'\frac{s}{2}\right)}, \tag{10}$$

[1]However, $\overline{\mathcal{V}}$ is disconnected and contains disjoint open intervals, due to the periodicity of the $\cos(\cdot)$ function.

which is well defined on $\overline{\mathcal{U}} \times \mathcal{S}$. Let $\mathcal{U} \subseteq \overline{\mathcal{U}}$ and $|\mathcal{U}|$ denote the number of points in $\mathcal{U}$, which can be chosen such that $|\mathcal{U}| \geq 1$. We introduce two simple minimum distance criterion functions on $\mathcal{S}$:[2]

$$J(s,\mathcal{U}) := \sum_{(u,u')\in\mathcal{U}} |\varphi_{\Delta p,2}(u,u') - \varphi_{\Delta p,1}(u)\varphi_{\Delta p,1}(u')R(u,u';s)|^2 \geq 0, \tag{11}$$

$$Q(s,\mathcal{U}) := \sum_{(u,u')\in\mathcal{U}} |H(u,u') - R(u,u';s)|^2 \geq 0, \tag{12}$$

where $|\cdot|$ denotes the modulus of a complex number. Since Eq. (7) (or (8)) holds for all $(u,u') \in \overline{\mathcal{V}}^2$ and $\mathcal{U} \subseteq \overline{\mathcal{U}} \subseteq \overline{\mathcal{V}}^2$, both criteria are minimized at $s = s_0$, i.e., $J(s_0,\mathcal{U}) = 0$ and $Q(s_0,\mathcal{U}) = 0$.

Assumption 2. *(i) $s_0 \in \mathcal{S}$, where $\mathcal{S}$ is compact; (ii) $\mathcal{U} \subseteq \overline{\mathcal{U}}$, and $\exists(\tilde{u},\tilde{u}) \in \mathcal{U}$ such that $\tilde{u} \in (0,\pi/\bar{s})$; and (iii) $|\mathcal{U}| < \infty$.*

As shown in Theorem 3 of Chen et al. (2017), under Assumption 1 and 2, s_0 is identified as the unique solution to $\min_{s\in\mathcal{S}} J(s,\mathcal{U})$ and $\min_{s\in\mathcal{S}} Q(s,\mathcal{U})$. For the identification of s_0 it suffices to choose a grid $\mathcal{U}$ satisfying Assumption 2(ii) with $|\mathcal{U}| = 1$. But a grid $\mathcal{U}$ with larger $|\mathcal{U}| > 1$ is better for more accurate estimation of s_0. Assumption 2(iii) is assumed for easy implementation of our simple estimators. Constructing $\mathcal{U}$ according to Section 2.1.1 will ensure that Assumption 2(ii) is satisfied with a grid $\mathcal{U}$ consisting of finitely many discrete points in $(0,\pi/\bar{s})^2 \cap \overline{\mathcal{V}}^2$.

Remark 1. Our model covers the case where the underlying true (log) price p_t^* has a possible drift. The observed price p_t then evolves according to

$$p_t = p_t^* + I_t\frac{s_0}{2}, \quad p_t^* = c_0 + p_{t-1}^* + e_t, \tag{13}$$

$$\Delta p_t = c_0 + e_t + \Delta I_t\frac{s_0}{2}. \tag{14}$$

Note that in this paper the distribution of ε_t is left completely unspecified, thus we could define $\varepsilon_t = c_0 + e_t$ (also applicable to the extended models). $\varphi_\varepsilon(\cdot)$ (the c.f. of ε_t) can be identified, see Chen et al. (2017) for details. Then c_0 could be identified as, for example, the mean of ε_t using $\varphi_\varepsilon(\cdot)$.

[2] If $|\mathcal{U}| = \infty$, there is a slight abuse of notations in definitions (11) and (12). Summations should be replaced by integrals with respect to some (positive) sigma-finite measure on $\mathcal{U}$.

We next introduce several simple spread estimators and then present their large sample properties.

2.1 New Simple Spread Estimators

Theorem 3 of Chen et al. (2017) suggests to estimate s_0 as a minimizer of the empirical version of the criterion (11) or (12). We first replace the population characteristic functions $\varphi_{\Delta p,2}$ and $\varphi_{\Delta p,1}$ by the corresponding empirical characteristic functions (e.c.f.), defined as

$$\varphi_{T,2}(u,u') = \frac{1}{T-1}\sum_{t=2}^{T} \exp\left(iu\Delta p_t + iu'\Delta p_{t-1}\right), \ \varphi_{T,1}(u) = \varphi_{T,2}(u,0) = \frac{1}{T}\sum_{t=1}^{T} \exp\left(iu\Delta p_t\right), \quad (15)$$

where $\{\Delta p_t\}_{t=1}^T$ denotes a sample of observed returns. Define $H_T(u,u') := \frac{\varphi_{T,2}(u,u')}{\varphi_{T,1}(u)\varphi_{T,1}(u')}$ as the empirical counterpart of $H(u,u')$. Two simple minimum distance estimators are then given by [3]

$$\widehat{s}_{ecf} := \operatorname*{arg\,min}_{s\in\mathcal{S}} \quad J_T\left(s,\mathcal{U}\right) = \sum_{(u,u')\in\mathcal{U}} |\varphi_{T,2}(u,u') - \varphi_{T,1}(u)\varphi_{T,1}(u')R(u,u';s)|^2, \qquad (16)$$

$$\widehat{s}_{ecf,2} := \operatorname*{arg\,min}_{s\in\mathcal{S}} \quad Q_T\left(s,\mathcal{U}\right) = \sum_{(u,u')\in\mathcal{U}} |H_T(u,u') - R(u,u';s)|^2. \qquad (17)$$

Both $\widehat{s}_{ecf}$ and $\widehat{s}_{ecf,2}$ belong to a class of minimum distance estimators. In the following, we provide a unified framework to analyze their properties. Let a grid $\mathcal{U}$ be such that $1 \leq |\mathcal{U}| < \infty$. Denote the vectorized versions of $\{H(u,u') : \forall(u,u') \in \mathcal{U}\}$, $\{H_T(u,u') : \forall(u,u') \in \mathcal{U}\}$ and $\{R(u,u';s) : \forall(u,u') \in \mathcal{U}\}$ as $H(\mathcal{U})$, $H_T(\mathcal{U})$ and $R(\mathcal{U};s)$, respectively. Let D be any positive semi-definite $|\mathcal{U}| \times |\mathcal{U}|$ matrix, which is conformable with the chosen grid vectorization. We define a general weighted minimum distance criterion

$$Q_D\left(s,\mathcal{U}\right) := \left[H(\mathcal{U}) - R(\mathcal{U};s)\right]^\intercal D\left[H(\mathcal{U}) - R(\mathcal{U};s)\right]. \qquad (18)$$

Note that $Q\left(s,\mathcal{U}\right) = Q_I\left(s,\mathcal{U}\right)$ and $J\left(s,\mathcal{U}\right) = Q_{D_0}\left(s,\mathcal{U}\right)$, where I is a $|\mathcal{U}| \times |\mathcal{U}|$ identity matrix and $D_0 = \operatorname{diag}\left\{|\varphi_{\Delta p,1}(u)|^2|\varphi_{\Delta p,1}(u')|^2 : \forall(u,u') \in \mathcal{U}\right\}$ conformable with the chosen grid vectorization.

[3] $\sum_{(u,u')\in\mathcal{U}} |H_T(u,u') - R(u,u';s)|^2 = \sum_{(u,u')\in\mathcal{U}} \left(\mathrm{Re}\left(H_T(u,u')\right) - R(u,u';s)\right)^2 + \sum_{(u,u')\in\mathcal{U}} \left(\mathrm{Im}\left(H_T(u,u')\right)\right)^2$, in which the second part does not depend on the parameter of interest.

A general weighted minimum distance estimator is then defined as follows:

$$\widehat{s}_{ecf,\widehat{D}_T} := \arg\min_{s\in\mathcal{S}} \quad Q_{\widehat{D}_T,T}(s,\mathcal{U}) = [\mathrm{Re}\,(H_T(\mathcal{U})) - R(\mathcal{U};s)]^\intercal\, \widehat{D}_T\,[\mathrm{Re}\,(H_T(\mathcal{U})) - R(\mathcal{U};s)]\,, \tag{19}$$

where $\widehat{D}_T$ is a consistent estimator of D. $\widehat{s}_{ecf,\widehat{D}_T}$ defines a class of minimum distance estimators, including $\widehat{s}_{ecf}$ and $\widehat{s}_{ecf,2}$. We show in Section 2.2 the $\sqrt{T}$ consistency and asymptotic normality of $\widehat{s}_{ecf,\widehat{D}_T}$. In principle, we can choose D to obtain the optimally weighted estimator $\widehat{s}^*_{ecf}$, i.e., the estimator that has the smallest asymptotic variance among the class of estimators (19).

For implementation, instead of using a numerical optimization routine to minimize the criteria $J_T(s,\mathcal{U})$, $Q_T(s,\mathcal{U})$, $Q_{\widehat{D}_T,T}(s,\mathcal{U})$ over the parameter space $\mathcal{S} = [0, \overline{s}]$, we apply a simple grid search over an equally spaced fine grid of $\mathcal{S}$. This is because simulations suggest that these criteria are only locally convex around s_0 and the numerical optimization might not work well (probably due to the periodicity of the involved $\cos(\cdot)$ functions in $R(\mathcal{U};s)$, see Figure A1 in Section A5 of the online supplement). And a grid search over $\mathcal{S}$ ensures that one picks the global minimum as the estimators.

2.1.1 Choice of a Grid $\mathcal{U}$

The choice of $\mathcal{U}$ plays an important role in the finite sample performance of our simple estimators, and therefore we discuss it in detail here. Due to the specific expressions of Eq. (7) or (8) and their empirical counterparts, it is sufficient and desirable to restrict the grid $\mathcal{U}$ consisting of points (u,u') close to the origin. To see this, suppose that the fundamental price innovations $\{\varepsilon_t\}$ have a density with respect to Lebesgue measure (which we do not assume, but also do not want to rule out). Since $\{\varepsilon_t\}$ and the increments of the trade direction indicators $\{\Delta I_t\}$ are independent by assumption, this implies that the observed price innovations $\{\Delta p_t\}$ have a density as well. The Riemann-Lebesgue lemma (see also Theorem 1.1.6 in Ushakov (1999)) implies that

$$\lim_{\|(u,u')\|\to\infty} \left|\varphi_{\Delta p,2}(u,u')\right| = 0. \tag{20}$$

But the e.c.f. $\varphi_{T,2}$ is the c.f. of a discrete distribution, and as such it is almost periodic (see, e.g., Exercise 1.8.6 in Bisgaard and Sasvári (2000)). Hence (see also Theorem 1.1.5 in Ushakov (1999)), regardless of the sample size T,

$$\limsup_{\|(u,u')\|\to\infty} \left|\varphi_{T,2}(u,u')\right| = 1. \tag{21}$$

This means that, at least for an absolutely continuous distribution of $\{\varepsilon_t\}$, the e.c.f. is not a good approximation of the true c.f. for large u, u'. Indeed, we find in simulations that the relative approximation error between the true c.f. and the e.c.f. increases exponentially with u, even for a large sample size (see Figure A2 in Section A5 of the online supplement). Thus, for large values of u, u', the moment conditions in (7) and (8) become very noisy, which appears to be problematic. This suggests to restrict $\mathcal{U}$ to points close to the origin to ensure that the e.c.f.'s are bounded away from zero by a certain magnitude. But the choice of $\mathcal{U}$ should depend on how fast the true c.f. $\varphi_{\Delta p,2}$ decays to zero, which is governed by the unknown distribution of ε_t and the unknown true spread s_0. To overcome this problem, we suggest the following data-driven construction of a suitable grid $\mathcal{U}$.

Algorithm:

(1) Compute the joint and marginal e.c.f.'s $\varphi_{T,2}(\cdot,\cdot)$ and $\varphi_{T,1}(\cdot)$ from the data.

(2) Choose a cutoff $c \in (0,1)$ and compute the largest value $\bar{u} \in (0, 0.95\,\pi/\bar{s}]$ for which

$$\min\left\{|\varphi_{T,2}(\bar{u},\bar{u})|, |\varphi^2_{T,1}(\bar{u})|\right\} \geq c.$$

We found in simulations that $c = 0.1$ works well; values of c close to 0 and 1 tend to increase the variance of the estimator.

(3) Choose a number $n_g \in \mathbb{N}$ and construct the grid $\mathcal{U} = \mathcal{V} \times \mathcal{V}$, where $\mathcal{V}$ contains n_g equally spaced points in $(0, \bar{u})$. We found in simulations that the accuracy of our simple estimators $\widehat{s}_{ecf}$ and $\widehat{s}_{ecf,2}$ turns to increase in the number of grid points; $n_g \geq 12$ seems to work well.

Remark 2. The above construction of $\mathcal{U}$ corresponds to trimming constraints $\mathcal{I}\left\{\left|\varphi^2_{T,1}(u)\right| \geq c\right\}$ and $\mathcal{I}\left\{|\varphi_{T,2}(u,u)| \geq c\right\}$. We show in the proof of Theorem 1, as long as the cutoff point c is chosen small enough, the trimming constraints are never binding asymptotically. $\square$

In addition to the proper choice of $\mathcal{U}$, another aspect of our estimation procedure also deserves attention. According to its definition in (7), the population quantity H satisfies $H(u, u') > 1$ for all small positive values u, u' whenever $s_0 > 0$. In finite samples, however, we often find that for the empirical counterpart H_T, its real part $\mathrm{Re}\,(H_T(u, u')) < 1$ for a number of points $(u, u') \in \mathcal{U}$, especially for small values of $s_0 > 0$ (for an illustration, see Figure A3 in Section A5 of the online supplement). This is simply due to sampling variation, and simulations confirm that the problem disappears with increasing sample size. This gives rise to the following problem: our estimation strategy minimizes the distance between $R(u, u'; s)$ and $H_T(u, u')$ over $\mathcal{S} = [0, \bar{s}]$. If $\mathrm{Re}\,(H_T(u, u')) < 1$, then $s = 0$ provides the "best fit" at (u, u'), in that $s = 0$ minimizes the distance between $R(u, u'; s)$ and $H_T(u, u')$, since $R(u, u'; s) > 1$ for $s > 0$ and $R(u, u'; 0) = 1$. If this happens for a large portion of the grid points, then the global minima of the empirical criterion functions Q_T, J_T and $Q_{\widehat{D}_T,T}$ will be shifted towards $s = 0$. However, such an estimate is not informative, although we encounter this phenomenon predominately for small samples and when the true s_0 is very close to zero. To avoid this downward bias, we suggest to exclude problematic grid points with $\mathrm{Re}\,(H_T(u, u')) < 1$ from the optimization step. This issue resembles the problem of a positive empirical covariance for the original Roll's estimator. However, instead of emulating the various proposals in the literature to deal with this issue (e.g., Hasbrouck (2009)'s suggestion to set the estimate to be 0 for a positive empirical covariance would correspond to setting $\mathrm{Re}\,(H_T(u, u')) = 1$), we simply remove the problematic points from the grid $\mathcal{U}$.

Remark 3. Instead of c.f.'s, we could use moment generating functions (m.g.f.'s). This would avoid the problem of singularities and periodicity, since all cosine functions would be replaced by the non-periodic and positive hyperbolic cosine functions. However, this comes at the cost of assuming $\{\varepsilon_t\}$ has a finite m.g.f. around the origin, which implies that all of its moments are finite. This is a strong assumption – in particular for finance applications – and goes against our desire to make minimal assumptions about the distribution of $\{\varepsilon_t\}$. We thus do not pursue this idea any further. $\square$

2.2 Large-Sample Properties of the Estimators

For any positive semi-definite weighting matrix D, and its consistent estimate $\widehat{D}_T$, we present the large sample properties of $\widehat{s}_{ecf,\widehat{D}_T}$ defined in (19). The class of estimators $\widehat{s}_{ecf,\widehat{D}_T}$ include $\widehat{s}_{ecf}$ ($D = D_0$) and $\widehat{s}_{ecf,2}$ ($D = I$) as special cases. The conditions are very weak.

Assumption 3. *(i) D is a positive semi-definite $|\mathcal{U}| \times |\mathcal{U}|$ matrix; and (ii) $\widehat{D}_T \to^p D$ as $T \to \infty$.*

Theorem 1. *Let Assumptions 1, 2 and 3 hold. Then: $\widehat{s}_{ecf,\widehat{D}_T} \to^p s_0$ as $T \to \infty$.*

In the following, ∇_s denotes the first derivative of a function with respect to s. Each component of $\nabla_s R(\mathcal{U}; s)$ is

$$\nabla_s R(u, u'; s) = \frac{\frac{u'}{2} \sin\left(u\frac{s}{2}\right) \cos\left(u\frac{s}{2}\right) + \frac{u}{2} \sin\left(u'\frac{s}{2}\right) \cos\left(u'\frac{s}{2}\right)}{\left[\cos\left(u\frac{s}{2}\right) \cos\left(u'\frac{s}{2}\right)\right]^2}. \tag{22}$$

Assumption 4. *(i) The true unknown s_0 lies in the interior of $\mathcal{S}$; and (ii) $\nabla_s R(\mathcal{U}; s_0)^\intercal D \nabla_s R(\mathcal{U}; s_0) > 0$.*

Theorem 2. *Suppose that Assumptions 1, 2, 3 and 4 hold. Then:*
(i) $\sqrt{T}\left(\widehat{s}_{ecf,\widehat{D}_T} - s_0\right) \to^d \mathcal{N}\left(0, Asyvar\left(\widehat{s}_{ecf,\widehat{D}_T}\right)\right)$, *with*

$$Asyvar\left(\widehat{s}_{ecf,\widehat{D}_T}\right) := \left(\nabla_s R(\mathcal{U}; s_0)^\intercal D \nabla_s R(\mathcal{U}; s_0)\right)^{-2} \times \nabla_s R(\mathcal{U}; s_0)^\intercal D \Sigma_0 D \nabla_s R(\mathcal{U}; s_0), \tag{23}$$

where Σ_0 is a positive definite $|\mathcal{U}| \times |\mathcal{U}|$ matrix defined in Section A2.1 of the online supplement ;
(ii) Based on (23), the optimally weighed estimator of s_0 is given by

$$\widehat{s}^*_{ecf} := \widehat{s}_{ecf,\widehat{\Sigma}_0^{-1}} = \arg\min_{s\in\mathcal{S}} Q_{\widehat{\Sigma}_0^{-1},T}(s, \mathcal{U}), \tag{24}$$

with $Asyvar\left(\widehat{s}^*_{ecf}\right) = \left(\nabla_s R(\mathcal{U}; s_0)^\intercal \Sigma_0^{-1} \nabla_s R(\mathcal{U}; s_0)\right)^{-1}$.

The asymptotic variances of all these estimators, $Asyvar(\widehat{s}_{ecf})$, $Asyvar(\widehat{s}_{ecf,2})$, $Asyvar\left(\widehat{s}_{ecf,\widehat{D}_T}\right)$ and $Asyvar\left(\widehat{s}^*_{ecf}\right)$, can be consistently estimated by replacing D_0, D, $\nabla_s R(\mathcal{U}; s_0)$ and Σ_0 by $\widehat{D}_0 = \text{diag}\left\{|\varphi_{T,1}(u)|^2 |\varphi_{T,1}(u')|^2 : \forall (u, u') \in \mathcal{U}\right\}$, $\widehat{D}_T$, $\nabla_s R(\mathcal{U}; \widehat{s})$ and $\widehat{\Sigma}_0$ respectively. Here $\widehat{s}$ is any consistent estimator of s_0 such as $\widehat{s}_{ecf}$ or $\widehat{s}_{ecf,2}$, and $\widehat{\Sigma}_0$ is a consistent estimator for Σ_0 given in Section A2.1 of the online supplement.

Remark 4. When $|\mathcal{U}| = 1$, i.e., the grid $\mathcal{U}$ consists of a single point (u, u) with $u \in (0, \pi/\bar{s}) \cap \overline{\mathcal{V}}$, our estimation procedure has a closed-form solution

$$\widehat{s}_{diag}(u) := \frac{2}{u}\text{arccos}\left(\sqrt{|H_T(u,u)|^{-1}}\right). \tag{25}$$

However, simulations suggest that the performance of our estimation procedure, in terms of RMSE, improves with $|\mathcal{U}|$ (the number of grid points). Nevertheless, averaging the estimates in (25) over various values of u could lead to efficiency gain. We leave this open for future research. □

Remark 5. One could drop Assumption 2(iii) to allow for infinitely many grid points (i.e., $|\mathcal{U}| = \infty$), and then apply an approach with a continuum of moment conditions similar to Carrasco et al. (2007). This alternative procedure could provide an asymptotically more efficient estimation of s_0 in theory. However, simulations indicate that it is computationally more demanding and no-clear efficiency gain in finite samples. Perhaps more importantly, the model is not first-order Markov and the semiparametric efficiency bound for s_0 is unknown. We leave it to future research for semiparametric efficient estimation of s_0. □

3 Adverse Selection and Large Sample Properties of Estimators

We now relax the basic Roll (1984) type model (1) to allow for adverse selection cost. Suppose that

$$p_t^* = p_{t-1}^* + \delta I_t + \varepsilon_t, \quad p_t = p_t^* + I_t\frac{s_0}{2}, \tag{26}$$

where δ measures the contribution of adverse selection (see, e.g., Glosten and Harris (1988), Huang and Stoll (1997), Neal and Wheatley (1998), Foucault et al. (2013)), i.e., the effect of a market order on the efficient price. It is believed that δ should be positive, since buyer initiated orders cause the underlying true prices to rise while seller initiated orders cause them to fall. Eq. (26) implies that

$$\Delta p_t = \varepsilon_t + \alpha_0 I_t - \beta_0 I_{t-1}, \tag{27}$$

where $\beta_0 = s_0/2$ and $\alpha_0 = s_0/2 + \delta$. Rewriting Eq.(27) in the form of our previous price dynamics in (2), i.e., $\Delta p_t = \tilde{\varepsilon}_t + (I_t - I_{t-1})s_0/2$, we have $\tilde{\varepsilon}_t = \varepsilon_t + \delta I_t$, and thus $\text{Cov}(\tilde{\varepsilon}_t, I_t) = \delta Var(I_t) \neq 0$, whenever $\delta \neq 0$. Therefore, the Roll and Hasbrouck estimators would be biased and inconsistent.

Using only information about the autocovariance of transaction prices, $(\alpha_0, \beta_0, \sigma_\varepsilon^2)$ cannot be jointly identified, even under Hasbrouck (2004)'s assumption of $\varepsilon_t \sim$ i.i.d. $\mathrm{N}(0, \sigma_\varepsilon^2)$. Section 5 of Chen et al. (2017) addresses the identification of the adverse selection model (27) with balanced order flow, with unbalanced order flow, and when $\{I_t\}$ has general discrete support. In this section, we consider the estimation of (α_0, β_0) assuming balanced order flow for simplicity, and establish the large sample properties of the proposed estimators. Theorems 3 and 4 can be extended to allow for unbalanced order flow and $\{I_t\}$ having general discrete support. This extension should be straightforward, but involves tedious calculation of the asymptotic variances.

Assumption 5. *(i) Data $\{p_t\}_{t=1}^T$ is generated from Eq. (27) with $\alpha_0,\ \beta_0 > 0$, where $\{\varepsilon_t\}$ is i.i.d. and independent of $\{I_t\}$; (ii) Assumption 1 (ii)(iii) holds.*

Assumption 5 implies $\{\Delta p_t\}_{t=1}^T$ is strictly stationary and 1-dependent. And for all $(u, u') \in \mathbb{R}^2$,

$$\varphi_{\Delta p,2}(u, u') = \varphi_\varepsilon(u)\varphi_\varepsilon(u') \cos(u\alpha_0) \cos\left(u'\alpha_0 - u\beta_0\right) \cos\left(u'\beta_0\right). \tag{28}$$

$$\textit{Denote} \quad \overline{\mathcal{U}}_{as} := \left\{(u, u') \in \overline{\mathcal{V}}^2 : \min_{(\alpha,\beta)\in\mathcal{S}^2} |\cos(u\beta)\cos\left(u'\alpha\right)| > 0\right\}, \tag{29}$$

and a function on $\overline{\mathcal{U}}_{as} \times \mathcal{S}^2$ as

$$R(u, u'; \alpha, \beta) := \frac{\cos(u'\alpha - u\beta)}{\cos(u\beta)\cos(u'\alpha)} = 1 + \tan(u\beta)\tan\left(u'\alpha\right).$$

Eq. (28) now implies that

$$H(u, u') = R(u, u'; \alpha_0, \beta_0) \quad for \ (u, u') \in \overline{\mathcal{V}}^2, \tag{30}$$

and hence $H(u, u')$ is real-valued for all $(u, u') \in \overline{\mathcal{V}}^2$. Since $\overline{\mathcal{V}}^2$ contains an open ball of $(0, 0)$, $\exists$ a small positive $\tilde{u} \in \overline{\mathcal{V}}$, such that $(\tilde{u}, \tilde{u}), (\tilde{u}, 2\tilde{u}), (2\tilde{u}, \tilde{u}) \in \overline{\mathcal{U}}_{as} \subset \overline{\mathcal{V}}^2$.

Assumption 6. *(i) $(\alpha_0, \beta_0) \in \mathcal{S}^2$, where $\mathcal{S} = [0, \overline{s}]$; (ii) $\mathcal{U} \subseteq \overline{\mathcal{U}}_{as}$ and $\exists (\tilde{u}, \tilde{u}), (\tilde{u}, 2\tilde{u}), (2\tilde{u}, \tilde{u}) \in \mathcal{U}$, such that $\tilde{u} \in \left(0, \frac{\pi}{2\overline{s}}\right)$; and (iii) $|\mathcal{U}| < \infty$.*

Theorem 7 of Chen et al. (2017) shows the identification result of (α_0, β_0). Denote the vectorized version of $\{R(u, u'; \alpha, \beta) : \forall (u, u') \in \mathcal{U}\}$ as $R(\mathcal{U}; \alpha, \beta)$. Similarly, for any positive semi-definite

$|\mathcal{U}| \times |\mathcal{U}|$ matrix D and its consistent estimator $\widehat{D}_T$, we can then define a general weighted minimum distance estimator as follows:

$$Q_{as,\widehat{D}_T,T}(\alpha,\beta,\mathcal{U}) := [\mathrm{Re}(H_T(\mathcal{U})) - R(\mathcal{U};\alpha,\beta)]^\intercal \widehat{D}_T [\mathrm{Re}(H_T(\mathcal{U})) - R(\mathcal{U};\alpha,\beta)],$$
$$\left(\widehat{\alpha}_{\widehat{D}_T}, \widehat{\beta}_{\widehat{D}_T}\right) := \underset{(\alpha,\beta)\in\mathcal{S}^2}{\arg\min} Q_{as,\widehat{D}_T,T}(\alpha,\beta,\mathcal{U}). \tag{31}$$

We could choose $D = D_0$ or D to be a $|\mathcal{U}| \times |\mathcal{U}|$ identity matrix for easy implementation. We now present the large sample properties of $\left(\widehat{\alpha}_{\widehat{D}_T}, \widehat{\beta}_{\widehat{D}_T}\right)$.

Theorem 3. [4] *Suppose that Assumptions 3, 5, and 6 hold. Then:* $\left(\widehat{\alpha}_{\widehat{D}_T}, \widehat{\beta}_{\widehat{D}_T}\right) \to^p (\alpha_0, \beta_0)$ *as* $T \to \infty$.

In the following, $\nabla_{(\alpha,\beta)}$ denotes the partial derivative of a function with respect to (α,β). $\nabla_{(\alpha,\beta)} R(\mathcal{U};\alpha,\beta)$ is a $|\mathcal{U}| \times 2$ matrix, each row of which is given by

$$\nabla_{(\alpha,\beta)} R(u,u';\alpha,\beta) = \left[\left(1+\tan^2(u'\alpha)\right) u' \tan(u\beta); \quad \left(1+\tan^2(u\beta)\right) u \tan(u'\alpha)\right]. \tag{32}$$

Assumption 7. *(i) The true unknown* (α_0,β_0) *lies in the interior of* $\mathcal{S}^2$*; and*
(ii) $\nabla_{(\alpha,\beta)} R(\mathcal{U};\alpha_0,\beta_0)^\intercal D \nabla_{(\alpha,\beta)} R(\mathcal{U};\alpha_0,\beta_0)$ *is nonsingular.*

Theorem 4. [5] *Suppose that Assumptions 3, 5, 6, and 7 hold. Then:*
$\sqrt{T}\begin{pmatrix} \widehat{\alpha}_{\widehat{D}_T} - \alpha_0 \\ \widehat{\beta}_{\widehat{D}_T} - \beta_0 \end{pmatrix} \to^d \mathcal{N}\left(0, Asyvar\left(\widehat{\alpha}_{\widehat{D}_T}, \widehat{\beta}_{\widehat{D}_T}\right)\right)$*, with* $Asyvar\left(\widehat{\alpha}_{\widehat{D}_T}, \widehat{\beta}_{\widehat{D}_T}\right) :=$

$$\left(\nabla_{(\alpha,\beta)} R(\mathcal{U};\alpha_0,\beta_0)^\intercal D \nabla_{(\alpha,\beta)} R(\mathcal{U};\alpha_0,\beta_0)\right)^{-1} \times \nabla_{(\alpha,\beta)} R(\mathcal{U};\alpha_0,\beta_0)^\intercal D \Omega_0 D \nabla_{(\alpha,\beta)} R(\mathcal{U};\alpha_0,\beta_0)$$
$$\times \left(\nabla_{(\alpha,\beta)} R(\mathcal{U};\alpha_0,\beta_0)^\intercal D \nabla_{(\alpha,\beta)} R(\mathcal{U};\alpha_0,\beta_0)\right)^{-1}, \tag{33}$$

where Ω_0 *is a positive definite* $|\mathcal{U}| \times |\mathcal{U}|$ *matrix defined in Section A2.2 of the online supplement.*

In principle, we can choose $D = \Omega_0^{-1}$ to obtain the optimally weighted estimator of (α_0,β_0).

[4] Given the identification results in Chen et al. (2017), Theorem 3 can be readily extended to the adverse selection model with unbalanced order flow and when $\{I_t\}$ has general discrete support.

[5] Theorem 4 and Section A2.2 of the online supplement provide an explicit formula of $Asyvar\left(\widehat{\alpha}_{\widehat{D}_T}, \widehat{\beta}_{\widehat{D}_T}\right)$ and a plug-in consistent estimator of it. The extension of Theorem 4 allowing for unbalanced order flow and $\{I_t\}$ having general discrete support is easy, but the calculations of the asymptotic variances are tedious. For future research, one might consider bootstrap methods.

4 Unbalanced Order Flow and Large Sample Properties of Estimators

Assumption 8. *(i) Assumption 1(i)(ii) holds; and (ii) $\{I_t\}$ takes values ± 1 with unknown probability $q_0 := \Pr(I_t = 1) \in (0,1)$.*

This relaxation allows for unbalanced order flow (i.e., $q_0 \neq 1/2$). If either $q_0 = 0$ or $q_0 = 1$, then $\Delta p_t = \varepsilon_t$ and the differenced data give no information about s_0, therefore we restrict $q_0 \in (0,1)$. Under Assumption 8, we obtain the following relations (similar to Eq.(4) in Section 2): for all $(u,u') \in \mathbb{R}^2$,

$$\begin{aligned}\varphi_{\Delta p,2}(u,u') = \varphi_\varepsilon(u)\varphi_\varepsilon(u') &\left[\cos\left(u\frac{s_0}{2}\right) + (2q_0-1)i\sin\left(u\frac{s_0}{2}\right)\right]\left[\cos\left(u'\frac{s_0}{2}\right) - (2q_0-1)i\sin\left(u'\frac{s_0}{2}\right)\right]\\ \times &\left[\cos\left((u'-u)\frac{s_0}{2}\right) + (2q_0-1)i\sin\left((u'-u)\frac{s_0}{2}\right)\right]. \end{aligned} \tag{34}$$

In addition to the definitions of $\overline{\mathcal{V}}$, $\overline{\mathcal{U}}$ and $H(u,u')$ given in Section 2, we introduce a function on $\overline{\mathcal{U}} \times \mathcal{S} \times (0,1)$ as

$$R(u,u';s,q) := \frac{\begin{array}{c}\left[\cos\left(u\frac{s}{2}\right) + (2q-1)i\sin\left(u\frac{s}{2}\right)\right]\left[\cos\left(u'\frac{s}{2}\right) - (2q-1)i\sin\left(u'\frac{s}{2}\right)\right] \\ \times\left[\cos\left((u'-u)\frac{s}{2}\right) + (2q-1)i\sin\left((u'-u)\frac{s}{2}\right)\right]\end{array}}{\left[\cos^2\left(u\frac{s}{2}\right) + (2q-1)^2\sin^2\left(u\frac{s}{2}\right)\right]\left[\cos^2\left(u'\frac{s}{2}\right) + (2q-1)^2\sin^2\left(u'\frac{s}{2}\right)\right]}. \tag{35}$$

In particular, $R(u,u';s,1/2) = R(u,u';s)$ as defined in Section 2. We have for all $(u,u') \in \overline{\mathcal{V}}^2$

$$H(u,u') = R(u,u';s_0,q_0) \Longleftrightarrow \begin{cases} \mathrm{Re}(H(u,u')) = \mathrm{Re}(R(u,u';s_0,q_0)) \\ \mathrm{Im}(H(u,u')) = \mathrm{Im}(R(u,u';s_0,q_0)) \end{cases}, \tag{36}$$

and $H(u,u')$ is complex-valued unless $q_0(q_0-1)(2q_0-1)\sin\left(u\frac{s_0}{2}\right)\sin\left(u'\frac{s_0}{2}\right)\sin\left((u'-u)\frac{s_0}{2}\right) = 0$. Since $\overline{\mathcal{V}}^2$ contains an open ball of $(0,0)$, $\exists$ a small positive $\tilde{u} \in \overline{\mathcal{V}}$, such that $(\tilde{u},\tilde{u}), (\tilde{u},-\tilde{u}) \in \overline{\mathcal{U}} \subset \overline{\mathcal{V}}^2$.

Assumption 9. *(i) $(s_0,q_0) \in \mathcal{S} \times [\underline{q},\overline{q}]$, where $\mathcal{S} = [0,\overline{s}]$ and $[\underline{q},\overline{q}] \subset (0,1)$; (ii) $\mathcal{U} \subseteq \overline{\mathcal{U}}$, and either (a) $\exists(\tilde{u},\tilde{u}), (\tilde{u},-\tilde{u}) \in \mathcal{U}$ or (b)[6] $\exists(\tilde{u},\tilde{u}), (\tilde{u},2\tilde{u})(2\tilde{u},2\tilde{u}) \in \mathcal{U}$, such that $\tilde{u} \in (0,\pi/\overline{s})$; and (iii) $|\mathcal{U}| < \infty$.*

[6] Note that $\frac{H(\tilde{u},2\tilde{u})}{H(\tilde{u},\tilde{u})H(2\tilde{u},2\tilde{u})} = \frac{H(\tilde{u},-\tilde{u})}{[H(\tilde{u},\tilde{u})]^2}$.

Assumption 10. *(i) D is a positive semi-definite $|2\mathcal{U}| \times |2\mathcal{U}|$ matrix; and (ii) $\widehat{D}_T \to^p D$ as $T \to \infty$.*

Theorem 4 of Chen et al. (2017) establishes the identification result of (s_0, q_0). Denote the vectorized version of $\{R(u, u'; s, q) : \forall (u, u') \in \mathcal{U}\}$ as $R(\mathcal{U}; s, q)$. Using Eq. (36), we can define a general weighted minimum distance estimator as follows:

$$Q_{un,\widehat{D}_T,T}(s, q; \mathcal{U}) := \begin{bmatrix} \operatorname{Re}\{H_T(\mathcal{U}) - R(\mathcal{U}; s, q)\} \\ \operatorname{Im}\{H_T(\mathcal{U}) - R(\mathcal{U}; s, q)\} \end{bmatrix}^\intercal \widehat{D}_T \begin{bmatrix} \operatorname{Re}\{H_T(\mathcal{U}) - R(\mathcal{U}; s, q)\} \\ \operatorname{Im}\{H_T(\mathcal{U}) - R(\mathcal{U}; s, q)\} \end{bmatrix},$$
$$\left(\widehat{s}_{\widehat{D}_T}, \widehat{q}_{\widehat{D}_T}\right) := \operatorname*{arg\,min}_{(s,q) \in \mathcal{S} \times [\underline{q}, \overline{q}]} Q_{un,\widehat{D}_T,T}(s, q; \mathcal{U}). \tag{37}$$

We could choose $D = \begin{bmatrix} D_0 & 0 \\ 0 & D_0 \end{bmatrix}$ or D to be a $|2\mathcal{U}| \times |2\mathcal{U}|$ identity matrix to simplify the estimation. The large sample properties of $\left(\widehat{s}_{\widehat{D}_T}, \widehat{q}_{\widehat{D}_T}\right)$ are presented in Theorems 5 and 6.

Theorem 5. *Suppose that Assumptions 8, 9, and 10 hold. Then: $\left(\widehat{s}_{\widehat{D}_T}, \widehat{q}_{\widehat{D}_T}\right) \to^p (s_0, q_0)$ as $T \to \infty$.*

Let $\nabla_{(s,q)}$ denote the partial derivative of a function with respect to (s, q). Denote the $|2\mathcal{U}|$ by 2 matrix

$$\nabla_{(s,q)} \underline{R}(\mathcal{U}; s, q)) = \left[\nabla_{(s,q)} \operatorname{Re}(R(\mathcal{U}; s, q))^\intercal, \nabla_{(s,q)} \operatorname{Im}(R(\mathcal{U}; s, q))^\intercal\right]^\intercal, \tag{38}$$

where $\nabla_{(s,q)} \operatorname{Re}(R(\mathcal{U}; s, q))$ and $\nabla_{(s,q)} \operatorname{Im}(R(\mathcal{U}; s, q))$ are $|\mathcal{U}|$ by 2 matrices, defined in Eq. (A11) of the online supplement.

Assumption 11. *(i) The true unknown (s_0, q_0) lies in the interior of $\mathcal{S} \times [\underline{q}, \overline{q}]$; and*
(ii) $\nabla_{(s,q)} \underline{R}(\mathcal{U}; s_0, q_0))^\intercal D \nabla_{(s,q)} \underline{R}(\mathcal{U}; s_0, q_0))$ is nonsingular.

Theorem 6. *Suppose that Assumptions 8, 9, 10 and 11 hold. Then:*
$\sqrt{T} \begin{pmatrix} \widehat{s}_{\widehat{D}_T} - s_0 \\ \widehat{q}_{\widehat{D}_T} - q_0 \end{pmatrix} \to^d \mathcal{N}\left(0, Asyvar\left(\widehat{s}_{\widehat{D}_T}, \widehat{q}_{\widehat{D}_T}\right)\right)$, with $Asyvar\left(\widehat{s}_{\widehat{D}_T}, \widehat{q}_{\widehat{D}_T}\right) :=$

$$\left(\nabla_{(s,q)} \underline{R}(\mathcal{U}; s_0, q_0))^\intercal D \nabla_{(s,q)} \underline{R}(\mathcal{U}; s_0, q_0))\right)^{-1} \times \nabla_{(s,q)} \underline{R}(\mathcal{U}; s_0, q_0))^\intercal D \Gamma_0 D \nabla_{(s,q)} \underline{R}(\mathcal{U}; s_0, q_0))$$
$$\times \left(\nabla_{(s,q)} \underline{R}(\mathcal{U}; s_0, q_0))^\intercal D \nabla_{(s,q)} \underline{R}(\mathcal{U}; s_0, q_0))\right)^{-1}, \tag{39}$$

where Γ_0 is a positive definite $|2\mathcal{U}| \times |2\mathcal{U}|$ matrix defined in Section A2.3 of the online supplement.

Theoretically, we can choose $D = \Gamma_0^{-1}$ to obtain the optimally weighted estimator of (s_0, q_0). *Asyvar* $\left(\widehat{s}_{\widehat{D}_T}, \widehat{q}_{\widehat{D}_T}\right)$ can be consistently estimated using a plug-in estimator. In practice, a more limited objective of detecting when order flow is unbalanced can be addressed by examining the imaginary part of $H(u, u')$ for $u \neq u'$ with small $u' \neq 0$, since for such cases, $H(u, u')$ is complex-valued when $q_0 \neq 1/2$ and is real-valued when $q_0 = 1/2$. This is what we implement in the empirical application Section 6.2.

5 Possibility of No Price Change and Large Sample Properties of Estimators

Assumption 12. *(i) Assumption 1(i)(ii) holds; and (ii) $\{I_t\}$ takes the value 0 with unknown probability $\pi_0 \in [0, 1)$ and takes values ± 1 with equal probability $\frac{1-\pi_0}{2}$.*

This relaxation reflects the fact that many transactions occur with no price change (see, e.g.,Huang and Stoll (1997)). The case of $\pi_0 = 1$ is ruled out, otherwise $\Delta p_t = \varepsilon_t$ and the differenced data give no information about s_0. Under Assumption 12, we obtain the following relations: for all $(u, u') \in \mathbb{R}^2$,

$$\varphi_{\Delta p,2}(u, u') = \varphi_\varepsilon(u)\varphi_\varepsilon(u') \left[\pi_0 + (1 - \pi_0) \cos\left(u\frac{s_0}{2}\right)\right] \times \left[\pi_0 + (1 - \pi_0) \cos\left(u'\frac{s_0}{2}\right)\right] \left[\pi_0 + (1 - \pi_0) \cos\left((u' - u)\frac{s_0}{2}\right)\right]. \tag{40}$$

In addition to $\overline{\mathcal{V}}$, $\overline{\mathcal{U}}$ and $H(u, u')$ defined in Section 2, we introduce a function on $\overline{\mathcal{U}} \times \mathcal{S} \times [0, 1)$ as

$$R(u, u'; s, \pi) := \frac{\pi + (1 - \pi) \cos\left((u' - u)\frac{s}{2}\right)}{\left[\pi + (1 - \pi) \cos\left(u\frac{s}{2}\right)\right] \left[\pi + (1 - \pi) \cos\left(u'\frac{s}{2}\right)\right]}, \tag{41}$$

which is real-valued. In particular, $R(u, u'; s, 0) = R(u, u'; s)$ defined in Section 2. We have :

$$H(u, u') = R(u, u'; s_0, \pi_0), \quad \textit{for all } (u, u') \in \overline{\mathcal{V}}^2. \tag{42}$$

Since $\overline{\mathcal{V}}^2$ contains an open ball of $(0,0)$, $\exists$ a small positive $\tilde{u} \in \overline{\mathcal{V}}$, such that $(\tilde{u},\tilde{u}),(\tilde{u},2\tilde{u}) \in \overline{\mathcal{U}} \subset \overline{\mathcal{V}}^2$. Let $\mathcal{U} \subseteq \overline{\mathcal{U}}$ which can be chosen to be $2 \leq |\mathcal{U}| < \infty$. Eq. (42) yields :

$$\pi_0 = \frac{2H(\tilde{u},\tilde{u})^{-1} - H(\tilde{u},2\tilde{u})^{-1} - 1}{4H(\tilde{u},\tilde{u})^{-1/2} - H(\tilde{u},2\tilde{u})^{-1} - 3}, \qquad \cos\left(\tilde{u}\frac{s_0}{2}\right) = \frac{H(\tilde{u},\tilde{u})^{-1/2} - \pi_0}{1-\pi_0}, \tag{43}$$

which can be used to identify (s_0, π_0). Section 3.2 of Chen et al. (2017) considers the case when $\{I_t\}$ may take values in $\{-k_1, \cdots, 0, \cdots, +k_2\}$ and gives more general identification result. Eq. (43) essentially provides a closed form solution to this basic no-price-change model. Theorems 7 and 8 can be extended to consider more general models. Although the extension is straightforward, the calculation of the asymptotic variances should be nontrivial.

Assumption 13. *(i)* $(s_0, \pi_0) \in \mathcal{S} \times [0, \overline{\pi}]$, *where* $\mathcal{S} = [0, \overline{s}]$ *and* $[0, \overline{\pi}] \subset [0, 1)$*; (ii)* $\mathcal{U} \subseteq \overline{\mathcal{U}}$ *and* $\exists(\tilde{u},\tilde{u}),(\tilde{u},2\tilde{u}) \in \mathcal{U}$*, such that* $\tilde{u} \in (0, \frac{\pi}{2\overline{s}})$*; and (iii)* $|\mathcal{U}| < \infty$.

Denote the vectorized version of $\{R(u,u';s,\pi) : \forall(u,u') \in \mathcal{U}\}$ as $R(\mathcal{U};s,\pi)$. For any positive semi-definite $|\mathcal{U}| \times |\mathcal{U}|$ matrix D and its consistent estimator $\widehat{D}_T$, we can define a general weighted minimum distance estimator as follows:

$$\begin{aligned} Q_{npc,\widehat{D}_T,T}(s,\pi;\mathcal{U}) &:= \left[\mathrm{Re}\left(H_T(\mathcal{U})\right) - R(\mathcal{U};s,\pi)\right]^\intercal \widehat{D}_T \left[\mathrm{Re}\left(H_T(\mathcal{U})\right) - R(\mathcal{U};s,\pi)\right], \\ \left(\widehat{s}_{\widehat{D}_T}, \widehat{\pi}_{\widehat{D}_T}\right) &:= \underset{(s,\pi)\in\mathcal{S}\times[0,\overline{\pi}]}{\arg\min}\, Q_{npc,\widehat{D}_T,T}(s,\pi;\mathcal{U}). \end{aligned} \tag{44}$$

D could be chosen as D_0 or a $|\mathcal{U}| \times |\mathcal{U}|$ identity matrix for easy implementation. We now present the large sample properties of $\left(\widehat{s}_{\widehat{D}_T}, \widehat{\pi}_{\widehat{D}_T}\right)$.

Theorem 7. *Suppose that Assumptions 3, 12 and 13 hold. Then :* $\left(\widehat{s}_{\widehat{D}_T}, \widehat{\pi}_{\widehat{D}_T}\right) \to^p (s_0, \pi_0)$ *as* $T \to \infty$.

Let $\nabla_{(s,\pi)}$ denote the partial derivative of a function with respect to (s,π). $\nabla_{(s,\pi)} R(\mathcal{U};s,\pi)$ is a $|\mathcal{U}|$ by 2 matrix and defined in Eq. (A15) of the online supplement.

Assumption 14. *(i) The true unknown* (s_0,π_0) *lies in the interior of* $\mathcal{S} \times [0,\overline{\pi}]$ *; and*
(ii) $\nabla_{(s,\pi)} R(\mathcal{U};s_0,\pi_0)^\intercal D \nabla_{(s,\pi)} R(\mathcal{U};s_0,\pi_0)$ *is nonsingular.*

Theorem 8. *Suppose that Assumptions 3, 12, 13 and 14 hold. Then:*
$$\sqrt{T}\begin{pmatrix} \widehat{s}_{\widehat{D}_T} - s_0 \\ \widehat{\pi}_{\widehat{D}_T} - \pi_0 \end{pmatrix} \to^d \mathcal{N}\left(0, Asyvar\left(\widehat{s}_{\widehat{D}_T}, \widehat{\pi}_{\widehat{D}_T}\right)\right), \textit{ with } Asyvar\left(\widehat{s}_{\widehat{D}_T}, \widehat{\pi}_{\widehat{D}_T}\right) :=$$

$$\begin{aligned} &\left(\nabla_{(s,\pi)}R(\mathcal{U};s_0,\pi_0)^\intercal D\nabla_{(s,\pi)}R(\mathcal{U};s_0,\pi_0)\right)^{-1} \times \nabla_{(s,\pi)}R(\mathcal{U};s_0,\pi_0)^\intercal D\Psi_0 D\nabla_{(s,\pi)}R(\mathcal{U};s_0,\pi_0) \\ &\qquad \times \left(\nabla_{(s,\pi)}R(\mathcal{U};s_0,\pi_0)^\intercal D\nabla_{(s,\pi)}R(\mathcal{U};s_0,\pi_0)\right)^{-1}, \end{aligned} \tag{45}$$

where Ψ_0 is a positive definite $|\mathcal{U}| \times |\mathcal{U}|$ matrix defined in Section A2.4 of the online supplement.

6 Simulation Studies and Empirical Application

We first present a simulation study that compares the finite sample performance of our estimators to the Roll (1984) serial covariance estimator and the Hasbrouck (2004) Gibbs sampling procedure. We then provide an empirical application to data on traded E-Mini S&P futures contracts for the day of the 2010 Flash Crash.

6.1 A Comparison of our Estimators to the Methods of Roll and Hasbrouck

We compare the finite sample performances of the following estimators : $\widehat{s}_{ecf}$ and $\widehat{s}_{ecf,2}$, which are based on the criteria J_T and Q_T, respectively; the "optimally" weighted estimator $\widehat{s}_{ecf,\widehat{\Sigma_0}^{-1}}$ defined in Eq.(24); and the estimators of Roll and Hasbrouck, denoted by $\widehat{s}_{Roll}$ and $\widehat{s}_{Has.}$, respectively. We use the following simulation designs:

- For the spread and the sample size, we follow Hasbrouck (2009) and use $s_0 \in \{0.02, 0.2\}$[7] and $T = 250$ (this corresponds to roughly one year of daily closing prices).

[7]Regarding the spread size, Hasbrouck (2009) notes the following ($c = s_0/2$ denotes the half-spread): "Although prior to 2000 the minimum price increment on most U.S. equities was \$0.125, it has since been \$0.01, and currently this value might well approximate the posted half-spread in a large, actively traded issue. For a share hypothetically priced at \$50, the implied c equals 0.0002. No approach using daily trade data is likely to achieve a precise estimate of such a magnitude. The posted half-spread for a thinly traded issue might be 25 cents on a \$5 stock, implying c equals 0.05. This is likely to be estimated much more precisely."

- For the distribution of ε_t, we consider five cases: $\varepsilon_t \sim 0.02 \times \mathrm{N}(0,1)$, as in Hasbrouck (2009); $\varepsilon_t \sim 0.02 \times t(1)$ and $\varepsilon_t \sim 0.02 \times t(2)$; as well as $\varepsilon_t \sim 0.02 \times \mathrm{LN}(0,1.25)$ and $\varepsilon_t \sim 0.02 \times \mathrm{LN}(0,2)$, where we center the log-normal (LN) distribution to have zero mean. For log prices, a standard deviation of 0.02 represents a daily volatility of 2%, and an annual volatility of about 32% (for 250 trading days).

- The number of simulation runs is $n = 5000$.

- For our estimators, we use the following parameters: $c = 0.1$, $n_g = 12$, and $\bar{s} = 0.05$ (for $s_0 = 0.02$) and $\bar{s} = 0.5$ (for $s_0 = 0.2$), along with 500 equally spaced points in $[0, \bar{s}]$ for $\mathcal{S}$. For $\widehat{s}_{ecf,\widehat{\Sigma_0}^{-1}}$ we use the regularized version $\left(\widehat{\Sigma_0} + 0.0001 \times I\right)^{-1}$ as the estimated weighting matrix.

- For the Roll's estimator, we use two versions: $\widehat{s}_{Roll,1}$ denotes the Roll's estimator with Hasbrouck (2009) correction (i.e., set the estimate to zero for a positive empirical covariance); and $\widehat{s}_{Roll,2}$ denotes the Roll's estimator with Harris (1990) correction (i.e., use the absolute value of the empirical covariance).

- For the Hasbrouck's estimator, we use the MATLAB code accompanying Hasbrouck (2004), provided on the author's website (retrieved on Oct 28, 2015), and use 10,000 sweeps of the Gibbs sampler with a burn-in of 2,000. We report two sets of results: $\widehat{s}_{Has.,1}$ denotes the Hasbrouck's estimator where we set the estimates to zero in case the procedure does not converge; $\widehat{s}_{Has.,n^*=\cdot}$ denotes the Hasbrouck's estimator where we only use the $n^* = \cdot$ simulation runs, out of $n = 5,000$, where the procedure converges.

The setup with Gaussian innovations represents a regime with thin tails, in which both Roll's and Hasbrouck's method should do well, given their embedded assumptions. The setup with heavy-tailed student-t innovations, however, should be challenging to those two methods, whereas we expect our estimators to be the more robust. The setup with (centered) log-normal innovations presents a regime with asymmetry, in which we expect Hasbrouck's estimator to be at a disadvantage. Indeed, these conjectures are confirmed in the simulation results, as presented in Tables 1 and 2. They can be summarized as follows:

- Our estimators $\widehat{s}_{ecf}$ and $\widehat{s}_{ecf,2}$ have very similar performance, with $\widehat{s}_{ecf}$ slightly better (in terms of RMSE) across all simulation designs. The optimally weighted estimator $\widehat{s}_{ecf,\widehat{\Sigma_0}^{-1}}$ does not work well in small samples ($T = 250$).

- Our estimators $\widehat{s}_{ecf}$ and $\widehat{s}_{ecf,2}$ are competitive in the thin-tailed regime, while both the Roll's and the Hasbrouck's method perform slightly better there. This is not surprising, given that those two methods are tailored to an environment with finite second moments. Moreover, the Hasbrouck's estimator is built on the assumption of normally distributed latent price innovations, which corresponds to the truth in this regime. However, the Hasbrouck's estimator is sensitive and may diverge frequently when the true unknown spread s_0 is large relative to the variance of the latent price innovation.[8]

- In the settings with student-t innovations, $\widehat{s}_{ecf}$ performs best. In particular, our estimator yields good results even in the extreme case of $\varepsilon_t \sim 0.02 \times t(1)$, where both Roll's and Hasbrouck's estimators do poorly, and our estimators beat those estimators by at least an order of magnitude in terms of RMSE. Although this case might be extreme, our empirical analysis in Section 6.2.2 suggests that it is not an unrealistic assumption for periods of heavy market turbulence . This makes the robustness of our estimator a relevant feature.

- In the asymmetric cases with $\varepsilon_t \sim 0.02 \times \mathrm{LN}(0, \cdot)$, our estimator $\widehat{s}_{ecf}$ again performs best.

[8]For example, the Hasbrouck's estimator only converges in about 60% out of $n = 5,000$ simulation runs when $s_0 = 0.2$ and $\varepsilon_t \sim 0.02 \times \mathrm{N}(0, 1)$. This is consistent with its behaviour in the empirical E-mini analysis: there, it does not converge because the price innovations seem to be discrete, up/down a tick; here, in the simulations, it also looks rather discrete, i.e., big (discrete) jumps of size $\pm s_0/2$, and comparably small variance of ε_t.

		RMSE	Bias	Stdev	$q_{2.5}$	q_{25}	q_{75}	$q_{97.5}$
$\varepsilon_t \sim 0.02\text{N}(0,1)$	$\widehat{s}_{ecf}$	0.0046	-0.0005	0.0046	0.0092	0.0167	0.0227	0.0274
	$\widehat{s}_{ecf,2}$	0.0051	-0.0007	0.0051	0.0085	0.0162	0.0229	0.0279
	$\widehat{s}_{ecf,\widehat{\Sigma_0}^{-1}}$	0.0110	**0.0001**	0.0110	0.0000	0.0150	0.0243	0.0492
	$\widehat{s}_{Roll,1}$	0.0042	-0.0003	0.0042	**0.0106**	**0.0173**	0.0225	0.0269
	$\widehat{s}_{Roll,2}$	**0.0041**	-0.0002	0.0041	**0.0106**	**0.0173**	0.0225	0.0269
	$\widehat{s}_{Has.,n^*=5000}$	0.0043	-0.0015	0.0041	0.0097	0.0157	**0.0215**	**0.0253**
$\varepsilon_t \sim 0.02t(2)$	$\widehat{s}_{ecf}$	**0.0053**	**0.0004**	**0.0053**	**0.0101**	**0.0167**	**0.0242**	0.0301
	$\widehat{s}_{ecf,2}$	0.0059	**0.0004**	0.0059	0.0097	0.0161	0.0247	0.0315
	$\widehat{s}_{ecf,\widehat{\Sigma_0}^{-1}}$	0.0187	-0.0087	0.0165	0.0000	0.0000	0.0157	0.0500
	$\widehat{s}_{Roll,1}$	0.0146	-0.0007	0.0146	0.0000	0.0048	0.0290	0.0469
	$\widehat{s}_{Roll,2}$	0.0123	0.0040	0.0116	0.0049	0.0163	0.0304	0.0499
	$\widehat{s}_{Has.,1}$	0.0087	-0.0073	0.0048	0.0086	0.0106	0.0137	**0.0209**
	$\widehat{s}_{Has.,n^*=4999}$	0.0087	-0.0073	0.0048	0.0086	0.0106	0.0137	**0.0209**
$\varepsilon_t \sim 0.02t(1)$	$\widehat{s}_{ecf}$	**0.0059**	0.0035	**0.0048**	0.0145	**0.0201**	**0.0268**	**0.0332**
	$\widehat{s}_{ecf,2}$	0.0063	**0.0031**	0.0055	0.0132	0.0192	0.0270	0.0341
	$\widehat{s}_{ecf,\widehat{\Sigma_0}^{-1}}$	0.0174	-0.0119	0.0127	0.0000	0.0000	0.0115	0.0500
	$\widehat{s}_{Roll,1}$	0.3816	0.1232	0.3612	0.0000	0.0000	0.1311	0.9691
	$\widehat{s}_{Roll,2}$	0.4088	0.1618	0.3754	0.0152	0.0527	0.1599	1.0288
	$\widehat{s}_{Has.,1}$	0.3493	0.1140	0.3302	**0.0203**	0.0339	0.1001	0.8295
	$\widehat{s}_{Has.,n^*=4948}$	0.3494	0.1141	0.3303	0.0205	0.0339	0.1001	0.8295
$\varepsilon_t \sim 0.02\text{LN}(0,1.25)$	$\widehat{s}_{ecf}$	**0.0040**	**0.0000**	0.0040	**0.0122**	**0.0174**	**0.0228**	0.0275
	$\widehat{s}_{ecf,2}$	0.0043	-0.0002	0.0043	0.0113	0.0169	**0.0228**	0.0277
	$\widehat{s}_{ecf,\widehat{\Sigma_0}^{-1}}$	0.0195	-0.0106	0.0163	0.0000	0.0000	0.0083	0.0500
	$\widehat{s}_{Roll,1}$	0.0190	0.0036	0.0187	0.0000	0.0000	0.0377	0.0573
	$\widehat{s}_{Roll,2}$	0.0196	0.0123	0.0152	0.0071	0.0217	0.0413	0.0641
	$\widehat{s}_{Has.,n^*=5000}$	0.0062	-0.0049	**0.0037**	0.0097	0.0128	0.0167	**0.0223**
$\varepsilon_t \sim 0.02\text{LN}(0,2)$	$\widehat{s}_{ecf}$	**0.0039**	0.0016	**0.0036**	**0.0146**	**0.0191**	0.0240	0.0286
	$\widehat{s}_{ecf,2}$	0.0040	**0.0010**	0.0039	0.0136	0.0183	**0.0237**	0.0285
	$\widehat{s}_{ecf,\widehat{\Sigma_0}^{-1}}$	0.0193	-0.0126	0.0147	0.0000	0.0000	0.0079	0.0500
	$\widehat{s}_{Roll,1}$	0.1718	0.1055	0.1356	0.0000	0.0000	0.1862	0.4265
	$\widehat{s}_{Roll,2}$	0.2163	0.1633	0.1419	0.0327	0.1018	0.2214	0.5458
	$\widehat{s}_{Has.,n^*=5000}$	0.1043	0.0685	0.0786	0.0323	0.0537	0.0979	0.2663

Table 1: Simulation results for spread $s_0 = 0.02$, sample size $T = 250$ and $n = 5,000$ simulation runs. In addition to simulation RMSE, Bias and Stdev, q_x is the $x\,\%$ quantile of the estimates across the simulation runs (an measure of dispersion of the estimators). $\widehat{s}_{ecf}$ and $\widehat{s}_{ecf,2}$ are our estimators based on criterion J_T and Q_T respectively; $\widehat{s}_{ecf,\widehat{\Sigma_0}^{-1}}$ is our "optimally" weighted estimator. $\widehat{s}_{Roll,1}$ and $\widehat{s}_{Roll,2}$ denote Roll's estimator with Hasbrouck (2009) correction and Harris (1990) correction respectively. $\widehat{s}_{Has.,n^*=\cdot}$ denotes Hasbrouck's estimator, where we only use the $n^* = \cdot$ simulation runs where the procedure converges. When $n^* < 5000$ we also report $\widehat{s}_{Has.,1}$, another Hasbrouck's estimator, where we set the estimate to zero in case the procedure does not converge.

		RMSE	Bias	Stdev	$q_{2.5}$	q_{25}	q_{75}	$q_{97.5}$
$\varepsilon_t \sim 0.02\mathrm{N}(0,1)$	$\widehat{s}_{ecf}$	0.0154	-0.0002	0.0154	0.1690	0.1900	0.2100	0.2300
	$\widehat{s}_{ecf,2}$	0.0156	-0.0002	0.0156	0.1680	0.1900	0.2100	0.2300
	$\widehat{s}_{ecf,\widehat{\Sigma_0}^{-1}}$	0.0528	0.0074	0.0523	0.1730	0.1920	0.2090	0.4345
	$\widehat{s}_{Roll,1}$	0.0143	0.0003	0.0143	0.1713	0.1910	0.2100	0.2283
	$\widehat{s}_{Roll,2}$	0.0143	0.0003	0.0143	0.1713	0.1910	0.2100	0.2283
	$\widehat{s}_{Has.,1}$	0.1292	-0.0836	0.0986	0.0000	0.0000	**0.2002**	**0.2031**
	$\widehat{s}_{Has.,n^*=2913}$	**0.0019**	**-0.0001**	**0.0019**	**0.1961**	**0.1986**	0.2011	0.2036
$\varepsilon_t \sim 0.02t(2)$	$\widehat{s}_{ecf}$	**0.0164**	**0.0000**	**0.0164**	**0.1670**	0.1890	0.2110	0.2320
	$\widehat{s}_{ecf,2}$	0.0166	**0.0000**	0.0166	**0.1670**	0.1890	0.2110	0.2320
	$\widehat{s}_{ecf,\widehat{\Sigma_0}^{-1}}$	0.0498	0.0059	0.0495	0.1630	0.1890	0.2120	0.3515
	$\widehat{s}_{Roll,1}$	0.0192	0.0006	0.0192	0.1661	0.1896	0.2115	0.2335
	$\widehat{s}_{Roll,2}$	0.0184	0.0008	0.0184	0.1663	0.1897	0.2115	0.2336
	$\widehat{s}_{Has.,1}$	0.0773	-0.0297	0.0714	0.0000	0.1967	**0.2048**	**0.2120**
	$\widehat{s}_{Has.,n^*=4343}$	0.0289	-0.0039	0.0286	0.0693	**0.1988**	0.2054	0.2123
$\varepsilon_t \sim 0.02t(1)$	$\widehat{s}_{ecf}$	**0.0186**	**-0.0009**	**0.0185**	**0.1610**	**0.1870**	**0.2120**	**0.2340**
	$\widehat{s}_{ecf,2}$	0.0187	-0.0010	0.0186	**0.1610**	0.1860	**0.2120**	**0.2340**
	$\widehat{s}_{ecf,\widehat{\Sigma_0}^{-1}}$	0.1031	0.0118	0.1025	0.0000	0.1470	0.2560	0.4970
	$\widehat{s}_{Roll,1}$	0.2397	0.0514	0.2342	0.0000	0.1700	0.2580	1.1315
	$\widehat{s}_{Roll,2}$	0.2410	0.0689	0.2310	0.0676	0.1772	0.2643	1.1652
	$\widehat{s}_{Has.,1}$	0.2973	-0.0398	0.2946	0.0523	0.0673	0.1405	0.7770
	$\widehat{s}_{Has.,n^*=4980}$	0.2976	-0.0392	0.2951	0.0530	0.0675	0.1408	0.7782
$\varepsilon_t \sim 0.02\mathrm{LN}(0,1.25)$	$\widehat{s}_{ecf}$	**0.0167**	**-0.0008**	**0.0167**	**0.1660**	0.1880	0.2110	0.2310
	$\widehat{s}_{ecf,2}$	0.0168	**-0.0008**	0.0168	0.1650	0.1880	0.2110	0.2310
	$\widehat{s}_{ecf,\widehat{\Sigma_0}^{-1}}$	0.0656	0.0115	0.0646	0.1505	0.1880	0.2140	0.4580
	$\widehat{s}_{Roll,1}$	0.0190	-0.0003	0.0190	0.1627	0.1881	0.2117	0.2354
	$\widehat{s}_{Roll,2}$	0.0188	-0.0003	0.0188	0.1627	0.1881	0.2117	0.2355
	$\widehat{s}_{Has.,1}$	0.0471	-0.0114	0.0457	0.0000	0.1987	**0.2087**	**0.2165**
	$\widehat{s}_{Has.,n^*=4870}$	0.0348	-0.0064	0.0343	0.0719	**0.1993**	0.2088	**0.2165**
$\varepsilon_t \sim 0.02\mathrm{LN}(0,2)$	$\widehat{s}_{ecf}$	**0.0214**	**-0.0017**	**0.0213**	**0.1550**	0.1840	**0.2130**	**0.2400**
	$\widehat{s}_{ecf,2}$	0.0215	**-0.0017**	0.0215	**0.1550**	0.1840	**0.2130**	**0.2400**
	$\widehat{s}_{ecf,\widehat{\Sigma_0}^{-1}}$	0.1383	-0.0023	0.1383	0.0000	0.0890	0.2720	0.5000
	$\widehat{s}_{Roll,1}$	0.1591	0.0253	0.1571	0.0000	0.1631	0.2844	0.5210
	$\widehat{s}_{Roll,2}$	0.1739	0.0591	0.1635	0.0672	**0.1847**	0.2961	0.5972
	$\widehat{s}_{Has.,n^*=5000}$	0.1380	-0.0916	0.1032	0.0569	0.0747	0.1095	0.2808

Table 2: Simulation results for spread $s_0 = 0.2$, sample size $T = 250$ and $n = 5,000$ simulation runs. (See the caption of Table 1 for further details.)

6.2 An Application to E-mini S&P Futures Transaction Data

In this section we apply our estimators to data on traded E-Mini S&P futures contracts. These contracts are electronically traded futures contracts with the S&P 500 stock market index as the underlying asset, where the notional value of each contract is 50 times the value of the S&P 500 index. The contracts are traded on the Chicago Mercantile Exchange's Globex electronic trading platform, where trading takes place from Sunday-Friday from 6 p.m. to 5 p.m. ET (Eastern Time), with a 15-minute trading halt period Monday-Friday from 4:15 p.m. to 4:30 p.m., and a maintenance period Monday-Thursday from 5 p.m. to 6 p.m..[9]

In our application, we look at the trading data for May 6, 2010.[10] During this day, financial markets in the U.S. experienced one of the most volatile periods on record, with major stock indices collapsing and rebounding within a short time frame of less than an hour.[11] Consequently, this episode has become known as the Flash Crash (of 2010). For an illustration, Figure 1 displays the transaction prices for the sample period: the left plot shows the trading price of the last trade in each second; the right plot shows the sequence of all transaction prices. The difference in the two plots highlights that the majority of the trades on May 6 happened around the time of the Flash Crash. For comparison purpose, Figure 2 displays the same data for May 13, 2010, on which no unusual market turbulence occurred. A joint report by the U.S. SEC and the U.S. CFTC (henceforth SEC-CFTC report) published in 2010 identifies the market for E-mini S&P futures as one of the sources of the turbulences: "The combined selling pressure from the sell algorithm, HFTs, and other traders drove the price of the E-Mini S&P 500 down approximately 3% in just four minutes from the beginning of 2:41 p.m. through the end of 2:44 p.m. During this same time cross-market arbitrageurs who did buy the E-Mini S&P 500, simultaneously sold equivalent amounts in the equities markets, driving the price of SPY (an exchange-Transaction fund which represents the S&P500 index) also down approximately 3%."

[9] Before September 21, 2015, E-mini contracts used to trade for 23 hours a day from 6 p.m. to 5:15 p.m. ET.

[10] Specifically, we look at all trades from 6 p.m., May 5 to 4:15 p.m., May 6 ET.

[11] For a more detailed description of the events on May 6, along with an in-depth empirical analysis, see, e.g., Kirilenko et al. (2014) or U.S. SEC & U.S. CFTC (2010).

This makes the E-mini futures market an interesting object to study. In particular, we want to analyze how the liquidity cost of the E-mini S&P future evolved during the period of the Flash Crash. We focus on the period from 2:32 p.m. to 3:08 p.m. ET (Kirilenko et al. (2014) date the Flash Crash to this specific period), and we restrict our analysis to trades in the E-mini contract maturing in June 2010 (this contract makes up 99.65% of the number of trades on that day). To measure the liquidity cost, we estimate the implied spread with our estimator $\widehat{s}_{ecf}$ (with $c = 0.1, n_g = 12$), as well as with $\widehat{s}_{Roll,1}$, i.e., the Roll's estimator with Hasbrouck (2009) correction. We do not report results for $\widehat{s}_{Has.}$, since the underlying Gibbs sampling procedure (with the parameter configurations as in the code of the author) only converged for about 20 % of the cases in the (restricted) sample. The method does not seem to handle high-frequency data well, which often involve consecutive trades at identical prices and price bounces in discrete (tick-size) steps. This makes the price innovations a discrete process, whereas the Hasbrouck's estimator is based on the assumption of Gaussian (and thus continuous) innovations. This is consistent with two observations: first, the convergent cases are concentrated around the most volatile subperiod, where the price innovations appear less discrete; and second, adding a small Gaussian noise to the data makes the algorithm converge. For estimation, we use a rolling-window approach, where we estimate the spread at each second, using all trades over the last 30 seconds as input data (alternative window sizes of 15 or 20 seconds do not change the results in a significant way). Figure 3 plots the corresponding prices and, at each second, the number of trades in the last 30 seconds for our restricted sample period. We use log prices to give the spread a relative percentage interpretation (given the magnitude, the results are restated in basis points, BPS; 1 BPS = 1/100 %).

The results using log prices are presented in Figures 4 and 5[12] and can be summarized as follows:

- Both estimators $\widehat{s}_{ecf}$ and $\widehat{s}_{Roll,1}$ produce almost identical (and roughly constant) results throughout the sample period, except for the time from 2:45 p.m. to 2:49 p.m. ET, during which the spread appears to spike, and then returns to its previous level. However, the increase is much more pronounced for $\widehat{s}_{Roll,1}$ than for our estimator $\widehat{s}_{ecf}$. The turbulence in

[12] Estimation results using level prices are presented in Figures A4 and A5 of the online supplement.

market prices during this period, along with the simulation evidence in the previous section on the robustness of $\widehat{s}_{ecf}$ in a heavy-tailed and asymmetric environment, suggest that $\widehat{s}_{Roll,1}$ might overstate the (increase in the) underlying liquidity cost, and that $\widehat{s}_{ecf}$ provides a better approximation. This is consistent with the fact that both methods produce nearly identical results outside the window of extreme turbulence.

- The detected spike in the spread is consistent with the following passages in the SEC-CFTC report: "HFTs, therefore, initially provided liquidity to the market. However, between 2:41 and 2:44 p.m., HFTs aggressively sold about 2,000 E-Mini contracts in order to reduce their temporary long positions." The estimates seem to pick up this temporary liquidity evaporation, although with some time lag.

- However, we do not find any detectable early warning signs of a pending crash in the spread estimates, which we will document below. This is in contrast to, e.g., Easley et al. (2012), who find that the (appropriately measured) market order flow became increasingly imbalanced in the hour preceding the crash, and that this imbalance contributed to the withdrawal of many liquidity providers from the market.

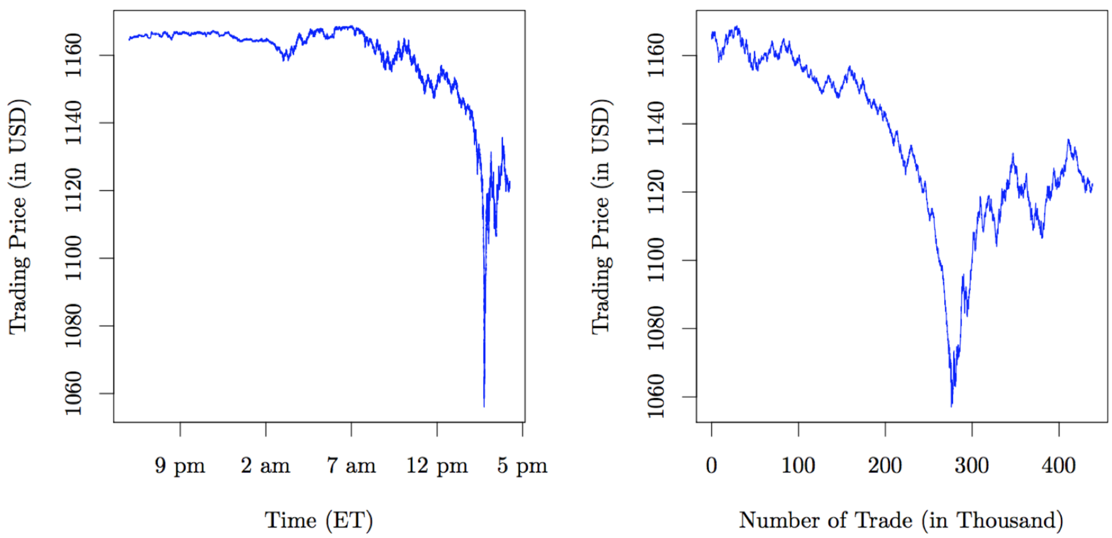

Figure 1: Transaction prices for E-Mini S&P futures (with maturity in June 2010) from 6 p.m. May 5, 2010 to 4:15 p.m. May 6, 2010, ET. Left: The last trading price for each second; Right: The sequence of all transaction prices throughout the day.

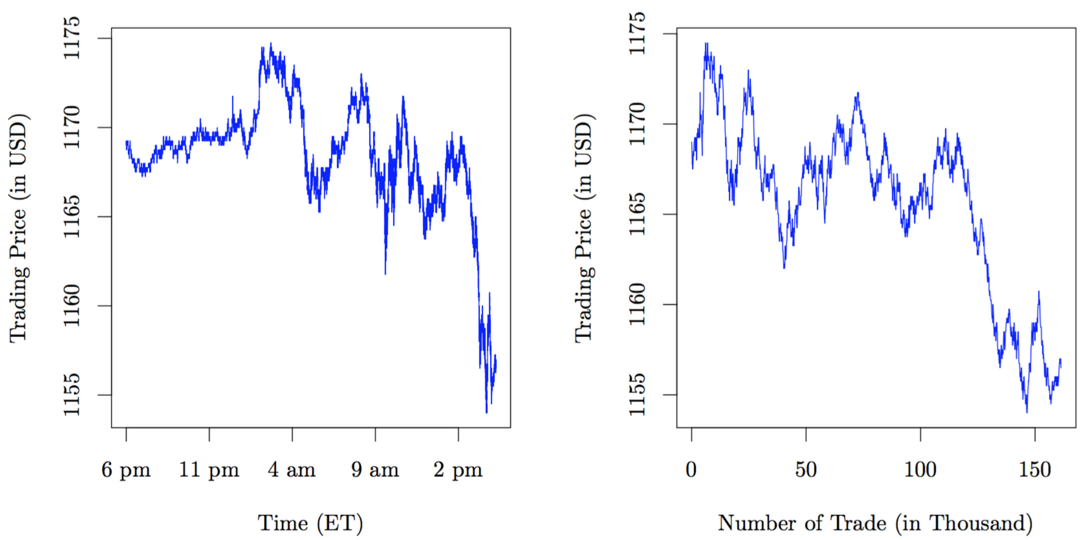

Figure 2: Transaction prices for E-Mini S&P futures (with maturity in June 2010) from May 12, 2010, 6 p.m. to May 13, 2010, 4:15 p.m. ET. Left: The last trading price for each second; Right: The sequence of all transaction prices throughout the day.

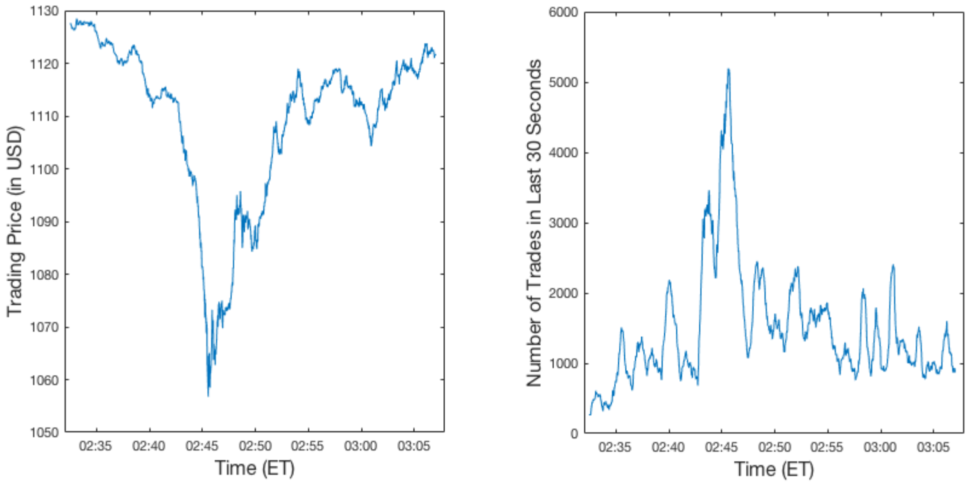

Figure 3: Transaction prices (left) and the number of trades in the last 30 seconds (right) for the period of the Flash Crash.

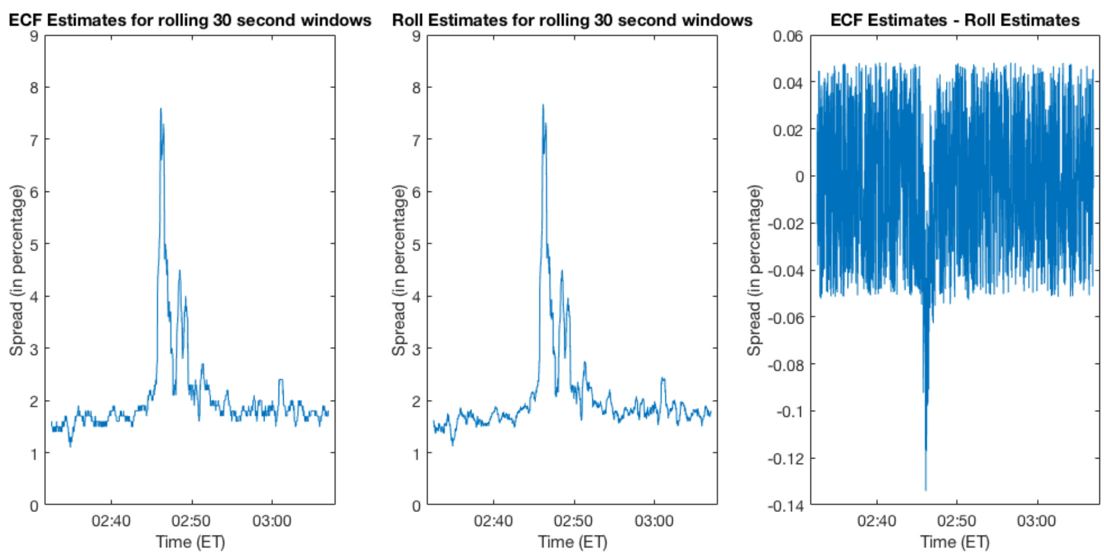

Figure 4: Spread estimates in percentage terms, for rolling 30 second windows (i.e., last 30 seconds of transactions are used for estimation) during the period of the Flash Crash.

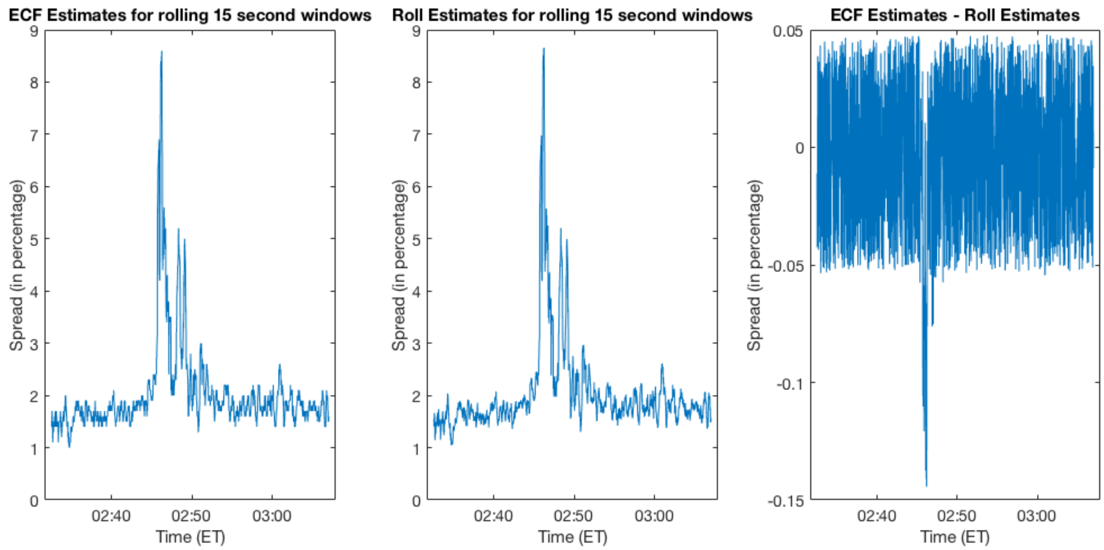

Figure 5: Spread estimates in percentage terms, for rolling 15 second windows (i.e., last 15 seconds of transactions are used for estimation) during the period of the Flash Crash.

6.2.1 Forecasting the Flash Crash

We estimate a bivariate VAR model for $(\Delta p, \log s)$ using the 107 observations of spreads (computed from non-overlapping thirty-second prior intervals) and the contemporaneous (log) prices. The results are presented in Table A1 of the online supplement (with standard errors in parentheses).

There is significant linear predictability in both series, although the spreads appear to be more predictable than returns according to the in-sample adjusted R-square measure.

The impulse response functions are shown in Figure A6 of the online supplement. They show that the return series is not significantly affected by shocks to the spread, whereas the spread does respond negatively to return shocks. The spread series is positively autocorrelated with a significant response (persistence) to past shocks.

The variance decompositions indicate that return variation is almost exclusively due to past returns, whereas spread variation is affected by past price especially after four lags. Despite this evidence of linear predictability in the two series over the whole period, linear models are not able, apparently, to forecast the largest movements during the peak period of the flash crash, as is shown by the residual graphs in Figure A7 of the online supplement.

6.2.2 Estimating the c.f. of the Fundamental Price Innovations ε_t

We have emphasized the estimation of the bid-ask spread parameter s_0, but it may also be of interest to estimate features of the distribution of the innovation process. We could obtain estimates of the c.f. of the innovation process directly from the data:

$$\widehat{\varphi}_\varepsilon(u) := \frac{\varphi_{T,2}(u,u)}{\varphi_{T,1}(u)}, \quad on\ \overline{\mathcal{V}}. \tag{46}$$

For an illustration, we estimate the c.f. φ_ε for three different points in time: before, at, and after the spike in the estimated spread (see Figures 4 and 5). Specifically, we choose the times 2:36 p.m., 2:46 p.m., and 2:56 p.m. ET, respectively. As in the previous section, we use all transaction prices for the last 30 seconds in the estimation. We find the following, with the estimates displayed in Figure 6:

- For 2:36 p.m., we obtain an estimate that resembles the c.f. of a point mass at zero (i.e., the real part is almost always equal to 1 and the imaginary part is very close to zero), which is intuitive. The data show that, during the tranquil periods of trading, the executed transaction price jumps up or down (with roughly equal probability) by at most a tick, which corresponds to $\varepsilon_t \approx 0$, i.e., there are no fundamental news, and the only price movements come from randomly arriving buy/sell orders.

- However, during the turbulent period, when the spread peaks at around 2:46 p.m., we obtain a significantly different behaviour of the latent price innovations. The estimate of $\mathrm{Re}(\varphi_\varepsilon(u))$ declines in a nearly linear fashion, which corresponds to the c.f. of a heavy-tailed distribution, while the estimate of $\mathrm{Im}(\varphi_\varepsilon(u))$ appears to be nonzero, which corresponds to the c.f. of an asymmetric distribution. This, again, is in line with economic intuition, since the crash in prices can be interpreted as reflection of a fundamental shock, represented by large and asymmetric innovations ε_t. In addition, this estimate supports our conjecture in Section 6.1 about the higher accuracy of our estimate $\widehat{s}_{ecf}$ compared to Roll's estimate $\widehat{s}_{Roll,1}$ during the turbulent period. Because based on the simulation evidence, the Roll's estimator performs worse under heavy-tailed and asymmetric innovations.

- After the peak turbulence, at 2:56 p.m., the estimate of the c.f. reflects a thin-tailed and symmetric regime again, close to the estimate that we obtain for 2:36 p.m..

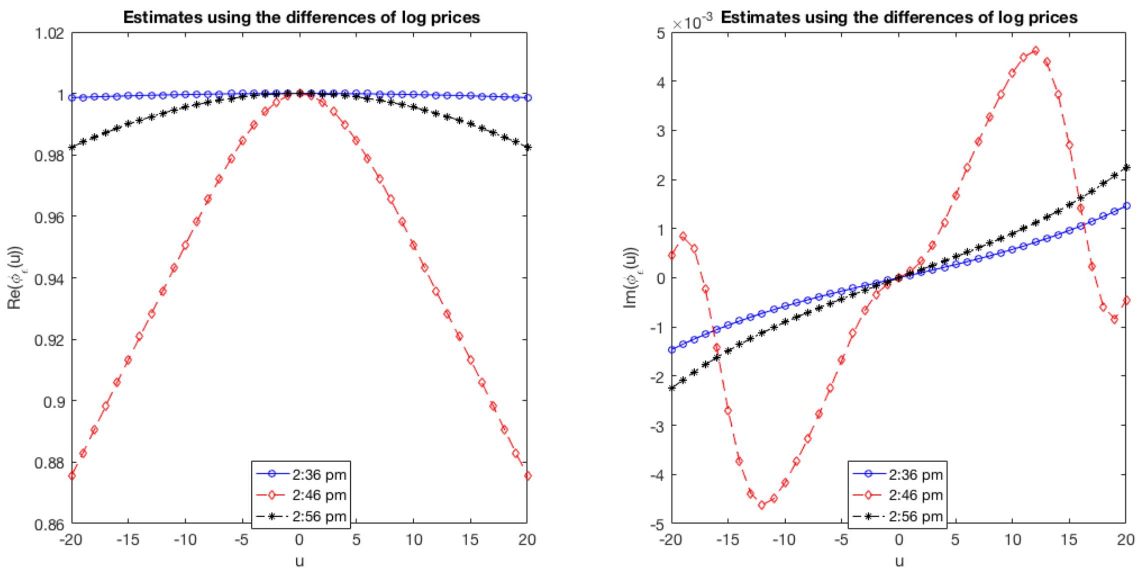

Figure 6: Estimates of the c.f. φ_ε of the latent price innovations ε_t during the Flash Crash of May 6, 2010. The plots/times refer to estimates before, at, and after the spike in the estimated spread, as displayed in Figures 4 and 5.

6.2.3 Detecting Order Flow Imbalances

Eq.(36) shows the following : under a balanced order flow (i.e., $I_t = \pm 1$ with equal probability), the population quantity $H(u, u')$ is real-valued, while under order flow imbalance (i.e., $\Pr(I_t = 1) \neq \Pr(I_t = -1)$), the quantity $H(u, u')$ is complex-valued when $u \neq u'$. This yields a way to detect order flow imbalances by measuring the imaginary part of the empirical quantity $H_T(u, u')$. In Figure 7, we plot the evolution of the two quantities

$$h_{max} := \max_{(u,u')\in\mathcal{U}} \left(\left|Im(H_T(u,u'))\right|\right) \qquad \text{and} \qquad h_{mean} := \frac{1}{|\mathcal{U}|} \sum_{(u,u')\in\mathcal{U}} \left(\left|Im(H_T(u,u'))\right|\right), \tag{47}$$

during the period of the Flash Crash. Clearly, the two measures h_{max} and h_{mean} spike during the peak turbulence (and are almost perfectly synchronized with the spread increase we detect). This indicates that not only the liquidity cost (as measured by the bid-ask spread) increases sharply, but also the order flow becomes highly imbalanced during this period. This is in line with the economic intuition of a panic sale interpretation of the crash.

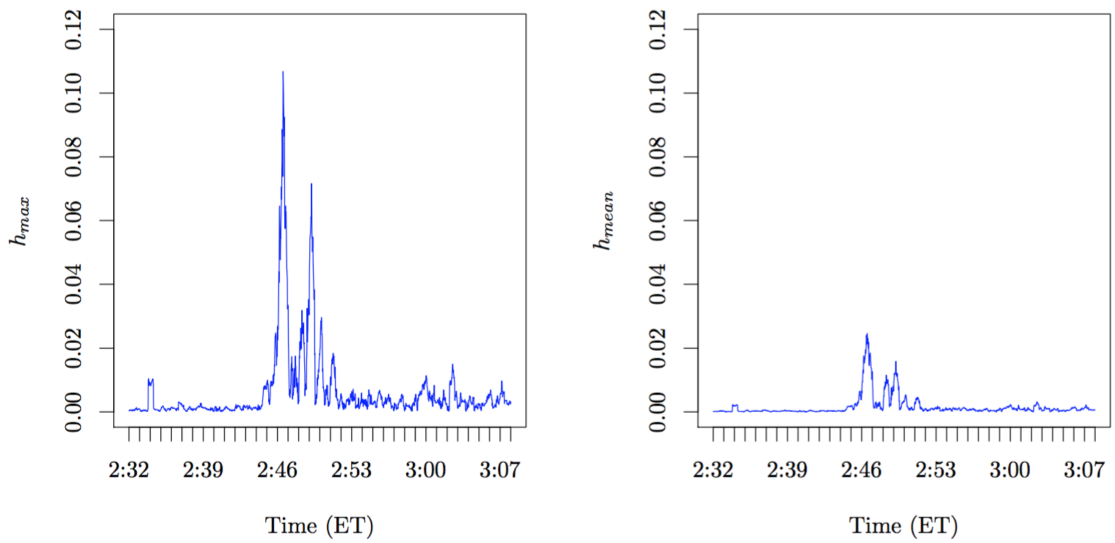

Figure 7: Indications of order flow imbalances during the Flash Crash of May 6, 2010. The definitions of the quantities h_{max} and h_{mean} are given in Eq.(47).

6.2.4 Aggregation Robustness

In Section A3 of the online supplement, we show that our estimators are aggregation robust. For the basic Roll (1984) model, we estimate the spread at each second, using the non-overlapping returns for every 2 ($k = 2$) and 5 ($k = 5$) transactions over the last 30 seconds. The results are presented in Figure A8 - A11 of the online supplement.

6.2.5 Adverse Selection Estimators

In Section 3, we consider the presence of an adverse selection component in the spread :

$$\Delta p_t = \varepsilon_t + \alpha_0 I_t - \beta_0 I_{t-1},$$

where $\beta_0 = s_0/2$ and $\alpha_0 = s_0/2 + \delta$. Using all trades over the last 15 and 30 seconds, we estimate $s = 2\beta$ and $\delta = \alpha - \beta$ at each second. Estimation results using log prices are presented in Figures 8 and 9[13]. The results show that in the run up to the Flash Crash, the adverse selection component of the spread was quite small. However, this rose substantially during the peak period.

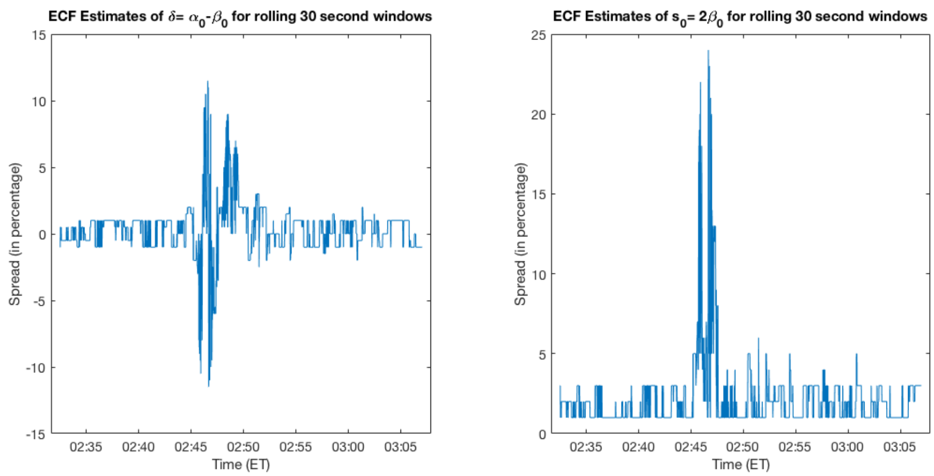

Figure 8: The adverse selection case : spread estimates in percentage terms, for rolling 30 second windows during the period of the Flash Crash.

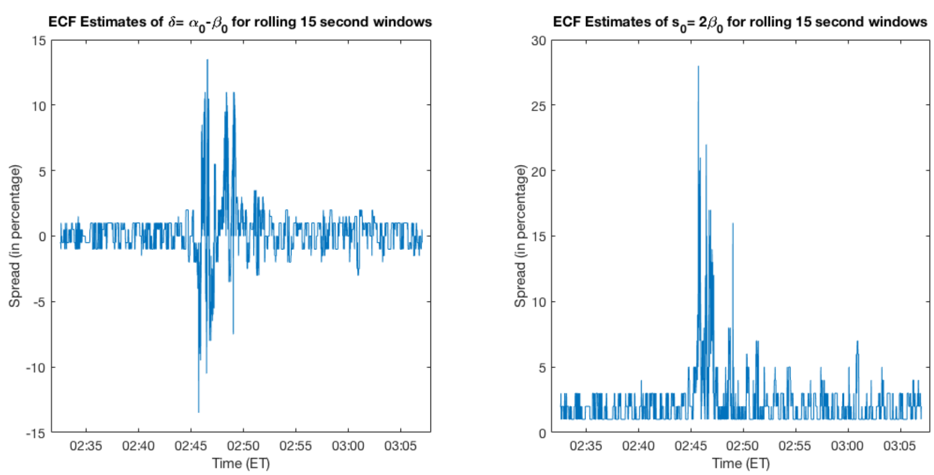

Figure 9: The adverse selection case : spread estimates in percentage terms, for rolling 15 second windows during the period of the Flash Crash.

[13]The estimation results using level prices are presented in Figures A12 and A13 of the online supplement.

7 Conclusions

In this paper we provide simple semiparametric estimators of the spread using transaction price data alone. We compare our method theoretically and numerically with the Roll (1984) estimator as well as with the Hasbrouck (2004) estimator. Our estimators perform similarly to theirs when the latent fundamental return distribution is Gaussian, but much better than theirs when the distribution is far from Gaussian, such as for heavy-tailed or asymmetric data.

Our c.f. based estimators are applied to the E-mini futures contract on the S&P 500 during the Flash Crash of 2010. We find that, during relatively tranquil times our estimator $\widehat{s}_{ecf}$ and the Roll estimator $\widehat{s}_{Roll,1}$ are very similar, while during the peak period of the Flash Crash, i.e., from 2:45 p.m. to 2:49 p.m. ET, the spread appears to spike, and then returns to its previous level, but the increase is much more pronounced for the Roll estimator than for our estimator. The estimated c.f. of ε_t indicates that the fundamental innovation becomes much more heavy-tailed and asymmetric during the turbulent period. Along with the simulation evidence on the robustness of our estimator $\widehat{s}_{ecf}$ in a heavy-tailed and asymmetric environment, it suggests that $\widehat{s}_{Roll}$ might overstate the underlying liquidity cost, and that $\widehat{s}_{ecf}$ provides a better approximation. This is consistent with the fact that both methods produce nearly identical results outside the window of extreme turbulence. We also find that the order flow becomes badly unbalanced and the adverse selection component of the spread fluctuates substantially during the peak period of the Flash Crash. Both of these findings corroborate the work presented in the SEC/CFTC report on the days events and subsequent academic work.

Acknowledgements

The authors would like to thank the guest coeditors, the two anonymous referees, Joel Hasbrouck, J. Huston McCulloch, and other seminar participants at Melbourne, NYU, Boston University, Lancaster, 2016 FERM and 2016 Shandong Econometrics Conference. Chen thanks Cowles Foundation for research support. Yi acknowledges supports from National Natural Science Foundation of China (Grant 71471107 and 71501120). Yi is also affiliated with the Key Laboratory of Mathematical Economics (SUFE), Ministry of Education.

References

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

Bisgaard, T. M. and Sasvári, Z. (2000). *Characteristic Functions and Moment Sequences: Positive Definiteness in Probability*. Nova Publishers.

Bleaney, M. and Li, Z. (2015). The performance of estimators of the bid-ask spread under less than ideal conditions. *Studies in Economics and Finance*, 32:98 – 127.

Brunnermeier, M. K. and Pedersen, L. H. (2005). Predatory trading. *Journal of Finance*, 60(4):1825–1863.

Carrasco, M., Chernov, M., Florens, J.-P., and Ghysels, E. (2007). Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics*, 140(2):529–573.

Chen, X., Linton, O., and Yi, Y. (2017). Semiparametric identification of the bid-ask spread in extended roll models. *Journal of Econometrics*, 200:312–325.

Corwin, S. and Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *Journal of Finance*, 67:719 – 760.

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.

Edwards, A., Harris, L. E., and Piwowar, M. (2007). Corporate bond market transaction costs and transparency. *Journal of Finance*, 62(3):1421–1451.

Foucault, T., Pagano, M., and Roell, A. (2013). *Market Liquidity: Theory, Evidence and Policy*. Oxford University Press.

Glosten, L. R. and Harris, L. E. (1988). Estimating the components of the bid/ask spread. *Journal of Financial Economics*, 21:123–142.

Glosten, L. R. and Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14:71–100.

Goyenko, R., Holden, C., and Trzcinka, C. (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics*, 92:153 – 181.

Harris, L. (1990). Statistical properties of the Roll serial covariance bid/ask spread estimator. *The Journal of Finance*, 45(2):579–590.

Harris, L. E. and Piwowar, M. (2006). Secondary trading costs in the municipal bond market. *Journal of Finance*, 61(3):1361–1397.

Hasbrouck, J. (2004). Liquidity in the futures pits: Inferring market dynamics from incomplete data. *Journal of Financial and Quantitative Analysis*, 39(02):305–326.

Hasbrouck, J. (2009). Trading costs and returns for us equities: Estimating effective costs from daily data. *The Journal of Finance*, 64(3):1445–1477.

Huang, R. D. and Stoll, H. R. (1997). The components of the bid-ask spread: A general approach. *The Review of Financial Studies*, 10(4):995–1034.

Ito, T. and Yamada, M. (2016). Puzzles in the forex tokyo "fixing": Order imbalances and biased pricing by banks. *NBER Working Paper No. 22820*.

Jankowitsch, R., Nashikkar, A., and Subrahmanyam, M. G. (2011). Price dispersion in otc markets: A new measure of liquidity. *Journal of Banking and Finance*, 35(2):343–357.

Kirilenko, A. A., Kyle, A. S., Samadi, M., and Tuzun, T. (2014). The flash crash: The impact of high frequency trading on an electronic market. *Available at SSRN 1686004*.

Lukacs, E. (1972). A survey of the theory of characteristic functions. *Advances in Applied Probability*, 4(1):1–38.

Muravyev, D. (2016). Order flow and expected option returns. *Journal of Finance*, 71:673–708.

Neal, R. and Wheatley, S. M. (1998). Adverse selection and bid-ask spreads : Evidence from closed-end funds. *Journal of Financial Markets*, 1:121–149.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance*, 39(4):1127–1139.

Schultz, P. (2001). Corporate bond trading costs: A peek behind the curtain. *Journal of Finance*, 56(2):677–698.

U.S. SEC & U.S. CFTC (2010). Findings regarding the market events of May 6, 2010 (Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues).

Ushakov, N. G. (1999). *Selected topics in characteristic functions*. Walter de Gruyter.