cemmap

centre for microdata methods and practice

Simple Nonparametric Estimators for the Bid-Ask Spread in the Roll Model

Xiaohong Chen
Oliver Linton
Stefan Schneeberger
Yanping Yi

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP12/16

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Simple Nonparametric Estimators for the Bid-Ask Spread in the Roll Model

Xiaohong Chen*
Yale University

Oliver Linton†
University of Cambridge

Stefan Schneeberger‡
Yale University

Yanping Yi §
Shanghai University of Finance and Economics

March 16, 2016

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 instead of the sample autocovariance function like the method of Roll (1984). As in Roll (1984), we have a closed form expression for the spread, but this is only based on a limited amount of the model-implied identification restrictions. We also provide methods that take account of more identification information. We compare our methods theoretically and numerically with the Roll method as well as with its best known competitor, the Hasbrouck (2004) method, which uses a Bayesian Gibbs methodology under a Gaussian assumption. Our estimators are competitive with Roll's and Hasbrouck's when the latent true fundamental return distribution is Gaussian, and 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. Extensions to

*Department of Economics, Yale University, PO Box 208281, New Haven CT 06520-8281, USA. E-mail: xiaohong.chen@yale.edu. Web Page: http://cowles.econ.yale.edu/faculty/chen.htm.

†Department of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom, E-mail: obl20@cam.ac.uk. Web Page: http://www.oliverlinton.me.uk.

‡Department of Economics, Yale University, New Haven CT 06520-8281, USA. E-mail: stefan.schneeberger@yale.edu.

§School of Economics, Shanghai University of Finance and Economics (SUFE), Shanghai, 200433, China. E-mail: yi.yanping@mail.shufe.edu.cn.

models allowing for unbalanced order flow or Hidden Markov trade direction indicators or trade direction indicators having general asymmetric sup port or adverse selection are also presented, without requiring additional data.

1 Introduction

The (quoted) bid-ask spread of a financial asset is the difference between the best quoted prices for an immediate purchase and an immediate sale of that asset. The spread represents a potential profit for the market maker handling the transaction, and is a major part of the transaction cost facing investors, especially since the elimination of commissions and the reduction in exchange fees that has happened in the last twenty years, see for example Jones (2002), Angel et al. (2011), and Castura et al. (2010). Measuring the bid ask spread in practice can be quite time consuming and subject to a number of potential accuracy issues due to the quoting strategies of High Frequency Traders, for example.

The seminal paper Roll (1984) provides a simple market microstructure model that allows one to estimate the bid-ask spread from observed transaction prices alone, without information on the underlying bid-ask price quotes and the order flow (i.e., whether a trade was buyer- or seller-induced). This is particularly useful for long historical data sets, which are often limited in their scope. For instance, Hasbrouck (2009) notes that "investigations into the role of liquidity and transaction costs in asset pricing must generally confront the fact that while many asset pricing tests make use of U.S. equity returns from 1926 onward, the high-frequency data used to estimate trading costs are usually not available prior to 1983. Accordingly, most studies either limit the sample to the post-1983 period of common coverage or use the longer historical sample with liquidity proxies estimated from daily data." Another area where the available data is limited are open-outcry markets (like the CME), in which bid and ask quotes by traders expire (if not filled) without recording (see, e.g., Hasbrouck (2004) for more details).

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

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

where p_t^* is the underlying fundamental (log) price with innovations ε_t, and the trade direction indicators $\{I_t\}$ are i.i.d. and take the values ± 1 with probability $q := \Pr(I_t = 1) = 1/2$, where

$I_t = 1$ indicates that the transaction is a purchase, and $I_t = -1$ denotes a sale. The price p_t is observed, whereas *all other variables in Equation* (1) *are unobserved.* The parameter of interest is the effective bid-ask spread s_0.[1] Roll (1984) assumes that $\{\varepsilon_t\}$ is serially uncorrelated and uncorrelated with the trade direction indicators $\{I_t\}$. Under these assumptions:

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

$$s_0 = 2\sqrt{-\operatorname{Cov}(\Delta p_t, \Delta p_{t-1})}. \tag{3}$$

Roll (1984) proposes to estimate s_0 from (3) by replacing the theoretical covariance by its empirical counterpart, i.e.,

$$\widehat{s}_{Roll} := 2\sqrt{-\widehat{\operatorname{Cov}}(\Delta p_t, \Delta p_{t-1})}. \tag{4}$$

In practice, this estimator is not satisfactory, since the empirical first-order autocovariance of price changes is often positive, in which case (4) 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 (4) 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.

In a well-known alternative, Hasbrouck (2004) proposes to strengthen Roll's modeling assumptions by assuming that $\{\varepsilon_t\}$ is i.i.d. with a known parametric distribution, and is independent of $\{I_t\}$.[2] He then uses a Bayesian Gibbs sampling methodology to estimate the spread parameter subject to a non-negativity constraint. Specifically, Hasbrouck (2004) assumes that $\varepsilon_t \sim$ i.i.d. $\mathrm{N}(0, \sigma_\varepsilon^2)$, where the parameter σ_ε is estimated jointly with the spread s_0. Corwin and Schultz (2012) propose another spread estimator based on consecutive daily high/low transaction prices. They also assume

[1] The bid-ask spread in Equation (1) is called effective bid-ask spread because it is based on the effective (average) price p_t that is paid to fill an order, and not necessarily on the quoted bid or ask price, since it might be the case that the order cannot be filled at the latter price (e.g., due to insufficient depth of the market).

[2] Hasbrouck (2004) presents an extension that relaxes the independence between $\{\varepsilon_t\}$ and $\{I_t\}$ assumption using additional trade volume data.

that the fundamental price process is a Geometric Brownian motion, which is even stronger than the discrete time Gaussian asumption employed in Hasbrouck (2004). The recent empirical literature emphasizes several issues with the Roll model. First, 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. Second, it assumes no serial correlation in trade direction indicators, i.e., I_t is uncorrelated with I_{t-j} for any $j \geq 1$. Third, market orders are assumed not to bring news into prices, so that I_t is uncorrelated with ε_{t+j} for $j \geq 0$. Fourth, expected returns are constant, which may be an unrealistic assumption for long horizon studies. Fifth, spreads themselves are constant within the sample period. Admitting any one of these effects in the model will lead to the undesired consequence that the spread estimators of Roll (1984) and Hasbrouck (2004) become *inconsistent* (i.e., biased even as sample size goes to infinity). Furthermore, without additional assumptions, or additional observed information, it may not be possible to identify the spread jointly with parameters describing order flow imbalance, for example.

There have been many recent suggestions for estimating spreads (and liquidity costs more generall), that relax some of these assumptions, but at the cost of requiring additional observed information (data) such as trade direction indicators. As we have said, 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) review these estimators and provide some comparison when the above assumptions, such as constant spread and i.i.d. mid-price increments, are not valid. 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 volume.

We work with the framework in (1), where only transaction prices are available. These prices could be daily or weekly closing prices, but might also consist of high-frequency intra-day prices. However, contrary to, e.g., Corwin and Schultz (2012), we do not require intra-day 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 $\{\Delta I_t\}$. The assumption of independence between $\{\varepsilon_t\}$ and $\{\Delta I_t\}$ 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. The consistency and asymptotic normality of our simple 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's and Hasbrouck's when the latent true fundamental return distribution is Gaussian, and perform much better when the distribution is either asymmetric or heavy-tailed. Since we are working with an independence assumption, we are also able to identify the characteristic function of the latent true fundamental price increments, as well as some further parameters associated with extensions to the basic Roll model. For example, parameters associated with unbalanced order flow and/or general asymmetric supported $\{I_t\}$, or those for Hidden Markov $\{I_t\}$, or those that capture an adverse selection component in the spread. Again, this can be accomplished without requiring additional data.

We apply our method 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, and indicators for an unbalanced order flow.

The rest of the paper is organized as follows: Section 2 presents the basic model and identification of the spread parameter. Section 3 provides new simple spread estimators and their asymptotic properties. Section 4 presents a simulation study and the empirical application. Section 5 considers extensions to the model that allow for unbalanced order flow, serially dependent latent trade indicator, or adverse selection. Section 6 concludes. All proofs are relegated to the Appendix.

2 Basic Model and Identification

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 Equation* (1) *with $s_0 > 0$, where $\{\varepsilon_t\}$ is i.i.d. and independent of $\{\Delta I_t\}$; (ii) $\{I_t\}$ is i.i.d.; and (iii) I_t takes the values ± 1 with equal probability.*

(See Section 5 for extended models by relaxing various parts of Assumption 1.) Let $\varphi_\varepsilon(u) :=$

$\mathbb{E}\left(\exp\left(iu\varepsilon_t\right)\right)$ denote the characteristic function (c.f.) of ε_t. Let $\varphi_{\Delta p,1}(u) := \mathbb{E}\left(\exp\left(iu\Delta p_t\right)\right)$ and $\varphi_{\Delta p,2}(u,u') := \mathbb{E}\left(\exp\left(iu\Delta p_t + iu'\Delta p_{t-1}\right)\right)$ denote the marginal and joint c.f. of Δp_t and $(\Delta p_t, \Delta p_{t-1})$, respectively. 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 $(u,u') \in \mathbb{R}^2$,

$$\begin{aligned}\varphi_{\Delta p,2}(u,u') &= \varphi_\varepsilon(u)\varphi_\varepsilon(u')\mathbb{E}\left(\exp\left(iu\Delta I_t\frac{s_0}{2} + iu'\Delta I_{t-1}\frac{s_0}{2}\right)\right)\\ &= \varphi_\varepsilon(u)\varphi_\varepsilon(u')\mathbb{E}\left(\exp\left(iuI_t\frac{s_0}{2}\right)\right)\mathbb{E}\left(\exp\left(i(u'-u)I_{t-1}\frac{s_0}{2}\right)\right)\mathbb{E}\left(\exp\left(-iu'I_{t-2}\frac{s_0}{2}\right)\right)\\ &= \varphi_\varepsilon(u)\varphi_\varepsilon(u')\cos\left(u\frac{s_0}{2}\right)\cos\left((u'-u)\frac{s_0}{2}\right)\cos\left(u'\frac{s_0}{2}\right). \qquad (5)\end{aligned}$$

Equation (5) evaluated at any $(u,0) \in \mathbb{R}^2$ yields the relation for the marginal c.f.:

$$\varphi_{\Delta p,1}(u) := \varphi_{\Delta p,2}(u,0) = \varphi_\varepsilon(u)\left[\cos\left(u\frac{s_0}{2}\right)\right]^2. \qquad (6)$$

Equation (5) evaluated at any $(u,u) \in \mathbb{R}^2$ yields another useful relation:

$$\varphi_{\Delta p,2}(u,u) = \left[\varphi_\varepsilon(u)\cos\left(u\frac{s_0}{2}\right)\right]^2. \qquad (7)$$

If the distribution of ε_t were parametrically specified, one could work directly with equations (5)-(7) 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. The multiplicative structure in (5), (6) and (7) reminds one of the proportional hazard model in Cox (1972), and we shall approach estimation in a similar way. We find a relation that eliminates the unknown function $\varphi_\varepsilon(\cdot)$, and then proceed to estimate the parametric model for the distribution of the trade direction effect. Denote

$$\overline{\mathcal{V}} := \left\{u \in \mathbb{R} : \varphi_{\Delta p,1}(u) \neq 0\right\}. \qquad (8)$$

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. Equations (6) and (7) imply that for all $u \in \overline{\mathcal{V}}$, $\varphi_{\Delta p,2}(u,u) \neq 0$, $\varphi_\varepsilon(u) \neq 0$ and $\cos\left(u\frac{s_0}{2}\right) \neq 0$ as well. We immediately obtain the following identification result.

Theorem 1. *Let Assumption 1 hold. Then: the c.f. $\varphi_\varepsilon(\cdot)$ is identified on $\overline{\mathcal{V}}$ as*

$$\varphi_\varepsilon(u) = \frac{\varphi_{\Delta p,2}(u,u)}{\varphi_{\Delta p,1}(u)}; \tag{9}$$

and the true spread s_0 is identified as

$$s_0 = \frac{2}{\tilde{u}} \arccos\left(\sqrt{\frac{\varphi^2_{\Delta p,1}(\tilde{u})}{\varphi_{\Delta p,2}(\tilde{u},\tilde{u})}}\right) \tag{10}$$

with a small positive $\tilde{u} \in \overline{\mathcal{V}}$.

Proof of Theorem 1: Under Assumption 1, Equations (6) and (7) hold for all $u \in \mathbb{R}$, which implies that for all $u \in \overline{\mathcal{V}}$, Equation (9) holds, and

$$\left|\cos\left(u\frac{s_0}{2}\right)\right| = \sqrt{\frac{\varphi^2_{\Delta p,1}(u)}{\varphi_{\Delta p,2}(u,u)}}, \tag{11}$$

which, at least for a small positive $\tilde{u} \in \overline{\mathcal{V}}$,[3] can be inverted to obtain Equation (10). □Once s_0 is identified, we may alternatively identify the c.f. $\varphi_\varepsilon(\cdot)$ using Equation (6) alone:

$$\varphi_\varepsilon(u) = \frac{\varphi_{\Delta p,1}(u)}{\left[\cos\left(u\frac{s_0}{2}\right)\right]^2}. \tag{12}$$

From (12) (or (9)), one may obtain cumulants of the noise distribution such as the variance by differentiating $\log \varphi_\varepsilon(u)$ at the origin.

2.1 Overidentification

The above closed-form identification result does not use all the model restrictions contained in Equation (5). We now present an alternative identification result (for s_0) that utilizes the fact that Equation (5) holds for all $(u,u') \in \mathbb{R}^2$. Denote

$$H(u,u') := \frac{\varphi_{\Delta p,2}(u,u')}{\varphi_{\Delta p,1}(u)\varphi_{\Delta p,1}(u')}, \tag{13}$$

which is well defined on $\overline{\mathcal{V}}^2$. Equations (5) and (6) imply that for all $(u,u') \in \overline{\mathcal{V}}^2$,

$$H(u,u') = \frac{\cos\left((u-u')\frac{s_0}{2}\right)}{\cos\left(u\frac{s_0}{2}\right)\cos\left(u'\frac{s_0}{2}\right)} =: R(u,u';s_0), \tag{14}$$

[3]Since $\cos(\cdot)$ is periodic and has countably many separated inverse values, it suffices to take a small positive $\tilde{u} \neq 0$.

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

$$\varphi_{\Delta p,2}(u,u') = \varphi_{\Delta p,1}(u)\varphi_{\Delta p,1}(u')R(u,u';s_0). \tag{15}$$

Equation (14) (or (15)) is free of the nuisance function $\varphi_\varepsilon(\cdot)$ and only depends on the parameter of interest s_0, which is the key insight of our identification and estimation methods. Equation (14) (or (15)) for identification of s_0 is similar to the classic GMM approach to identification and 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 Equation (14) (or (15)) contains infinitely many overidentifying restrictions for s_0. Let $\mathcal{S} := [0,\overline{s}]$ denote the parameter space, where $\overline{s} > 0$ is chosen from prior experience for the market (to ensure that $s_0 \in \mathcal{S}$). Denote

$$\overline{\mathcal{U}} := \left\{(u,u') \in \overline{\mathcal{V}}^2 : \min_{s\in\mathcal{S}} |\cos\left(u\frac{s}{2}\right)\cos\left(u'\frac{s}{2}\right)| > 0\right\}, \tag{16}$$

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

$$R(u,u';s) := \frac{\cos\left((u-u')\frac{s}{2}\right)}{\cos\left(u\frac{s}{2}\right)\cos\left(u'\frac{s}{2}\right)}, \tag{17}$$

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 to be $|\mathcal{U}| \geq 1$. We introduce two simple minimum distance criterion functions on $\mathcal{S}$:[4]

$$J(s,\mathcal{U}) := \sum_{(u,u')\in\mathcal{U}} |\varphi_{\Delta p,2}(u,u') - \varphi_{\Delta p,1}(u)\varphi_{\Delta p,1}(u')R(u,u';s)|^2 \geq 0, \tag{18}$$

$$Q(s,\mathcal{U}) := \sum_{(u,u')\in\mathcal{U}} |H(u,u') - R(u,u';s)|^2 \geq 0. \tag{19}$$

Since Equation (14) (or (15)) 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) $s_0 \in \mathcal{S}$; (ii) either (a) $\mathcal{U} = \overline{\mathcal{U}}$; or (b) $\mathcal{U} \subset \overline{\mathcal{U}}$, and $\exists(\tilde{u},\tilde{u}) \in \mathcal{U}$ such that $0 < \tilde{u} < \overline{u}$, where $\overline{u}$ denotes the first positive zero of $u \mapsto \min_{s\in\mathcal{S}} \cos\left(u\frac{s}{2}\right)$.*

Theorem 2. *Let Assumptions 1 and 2 hold. Then: s_0 is identified as the unique solution to $\min_{s\in\mathcal{S}} J(s,\mathcal{U})$ or to $\min_{s\in\mathcal{S}} Q(s,\mathcal{U})$, and satisfies the identifiable uniqueness on $\mathcal{S}$.*[5]

[4] If $|\mathcal{U}| = \infty$, there is a slight abuse of notations in definitions (18) and (19). Summations should be replaced by integrals with respect to some (positive) sigma-finite measure on $\mathcal{U}$.

[5] That is, for all sequences $\{a_k\} \subset \mathcal{S}$ with $J(a_k,\mathcal{U})$ (or $Q(a_k,\mathcal{U})$) going to 0, we have $|a_k - s_0|$ goes to zero.

We do not impose any restriction on the error distribution.

Assumptions 1 and 2 are sufficient for the identification of s_0. Constructing $\mathcal{U}$ according to Section 3.1.1 will ensure that Assumption 2(ii)(b) is satisfied with a grid $\mathcal{U}$ consisting of finitely many discrete points in $(0, \overline{u})^2$.

As shown in Theorem 1, for the identification of s_0 it suffices to choose a grid $\mathcal{U}$ satisfying Assumption 2(ii)(b) with $|\mathcal{U}| = 1$. But a grid $\mathcal{U}$ with larger $|\mathcal{U}| > 1$ is better for more accurate estimation of s_0. Theorem 2 suggests a natural minimum distance estimation procedure for s_0 in Section 3.

3 Estimators and Asymptotic Properties

This section introduces several simple spread estimators and then presents their large sample properties.

3.1 New Simple Spread Estimators

The identification Theorem 2 suggests to estimate s_0 as a minimizer of the empirical version of the criterion (18) or (19). 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

$$\varphi_{T,2}(u, u') = \frac{1}{T-1} \sum_{t=2}^{T} \exp\left(iu\Delta p_t + iu'\Delta p_{t-1}\right), \tag{20}$$

$$\varphi_{T,1}(u) := \varphi_{T,2}(u, 0) = \frac{1}{T} \sum_{t=1}^{T} \exp\left(iu\Delta p_t\right), \tag{21}$$

where $\{\Delta p_t\}_{t=1}^{T}$ denotes a sample of observed price changes. Define

$$H_T(u, u') := \frac{\varphi_{T,2}(u, u')}{\varphi_{T,1}(u)\varphi_{T,1}(u')}$$

as the empirical counterpart of $H(u, u')$. The empirical criterion functions are simply given by

$$J_T(s, \mathcal{U}) := \sum_{(u,u') \in \mathcal{U}} |\varphi_{T,2}(u, u') - \varphi_{T,1}(u)\varphi_{T,1}(u')R(u, u'; s)|^2, \tag{22}$$

$$Q_T(s, \mathcal{U}) := \sum_{(u,u') \in \mathcal{U}} |H_T(u, u') - R(u, u'; s)|^2. \tag{23}$$

We use the absolute value in (22) and (23) to obtain a real-valued criterion that we can optimize. The estimators $\widehat{s}_{ecf}$ and $\widehat{s}_{ecf,2}$ solve

$$\widehat{s}_{ecf} := \arg\min_{s\in\mathcal{S}} J_T\left(s,\mathcal{U}\right), \tag{24}$$

$$\widehat{s}_{ecf,2} := \arg\min_{s\in\mathcal{S}} Q_T\left(s,\mathcal{U}\right). \tag{25}$$

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. Denote

$$D_0 = \operatorname{diag}\left\{|\varphi_{\Delta p,1}(u)|^2|\varphi_{\Delta p,1}(u')|^2 : \forall(u,u') \in \mathcal{U}\right\} \tag{26}$$

conformable with the chosen grid vectorization. For any positive semi-definite $|\mathcal{U}| \times |\mathcal{U}|$ matrix D, we can define a general weighted minimum distance criterion

$$Q_D\left(s,\mathcal{U}\right) := \left[H(\mathcal{U}) - R(\mathcal{U};s)\right]^{\intercal} D\left[H(\mathcal{U}) - R(\mathcal{U};s)\right], \tag{27}$$

which include the criteria (18) and (19) as special cases: $Q_{D_0}\left(s,\mathcal{U}\right) = J\left(s,\mathcal{U}\right)$ and $Q_I\left(s,\mathcal{U}\right) = Q\left(s,\mathcal{U}\right)$. We can define a general weighted minimum distance estimator as follows:

$$\begin{aligned} Q_{\widehat{D}_T,T}\left(s,\mathcal{U}\right) &:= \left[Re\left(H_T(\mathcal{U})\right) - R(\mathcal{U};s)\right]^{\intercal} \widehat{D}_T\left[Re\left(H_T(\mathcal{U})\right) - R(\mathcal{U};s)\right], \\ \widehat{s}_{ecf,\widehat{D}_T} &:= \arg\min_{s\in\mathcal{S}} Q_{\widehat{D}_T,T}\left(s,\mathcal{U}\right), \end{aligned} \tag{28}$$

where $\widehat{D}_T$ is a consistent estimator of D. We show in Section 3.2.2 how to 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 minimum distance estimators (28).

In computation, instead of using a numerical optimization routine to minimize the criteria $J_T\left(s,\mathcal{U}\right)$, $Q_T\left(s,\mathcal{U}\right)$, $Q_{\widehat{D}_T,T}\left(s,\mathcal{U}\right)$ over the parameter space $\mathcal{S} = [0,\bar{s}]$, we apply a simple grid search over an equally spaced fine grid of $\mathcal{S}$. This is because simulations suggest that these criteria might exhibit many local minima (due to the periodicity of the involved $\cos(\cdot)$ functions in $R(\mathcal{U};s)$), and a grid search over $\mathcal{S}$ ensures that one picks the global minimum as the estimators.

3.1.1 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 so we discuss it in some detail here. Due to the specific expressions of Equation (14) or (15)

and their empirical counter parts, 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 ε_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 Δp have a density as well. The Riemann-Lebesgue lemma (see also Theorem 1.1.6 in Ushakov (1999)) implies that

$$\lim_{\|(u,u')\|\to\infty} \left|\varphi_{\Delta p,2}(u, u')\right| = 0. \tag{29}$$

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,

$$\limsup_{\|(u,u')\|\to\infty} \left|\varphi_{T,2}(u, u')\right| = 1. \tag{30}$$

This means that, at least for an absolutely continuous distribution of ε_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 8 in Appendix D.). Thus, for large values of u, u', the moment conditions in (14) and (15) 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 how close to the origin such points should be depends on how fast the true c.f. $\varphi_{\Delta p,2}$ decays to zero, which in turn is governed by the distribution of ε_t and the true spread s_0, both of which are unknown. 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

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

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 $\widehat{s}_{ecf,2}$ turn to increase in the number of grid points; $n_g \geq 12$ seems to work well.

Remark 1. The above construction of $\mathcal{U}$ correspond to trimming constraints $\mathcal{I}\left\{|\varphi_{T,1}^2(u)| > c\right\}$ and $\mathcal{I}\left\{|\varphi_{T,2}(u,u)| > c\right\}$. We show in the proof of Theorem 3 that, as long as the cutoff point c is chosen small enough, the trimming constraints are never binding asymptotically. □

In addition to the proper choice of $\mathcal{U}$, there is another aspect of our estimation procedure that deserves attention. According to its definition in (14), 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 $\mathrm{Re}\left(H_T(u,u')\right) < 1$ for a number of the points $(u,u') \in \mathcal{U}$, especially for small values of $s_0 > 0$ (for an illustration see Figure 7 in Appendix D.). 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 $\mathrm{Re}\left(H_T(u,u')\right) < 1$, then $s = 0$ provides the "best fit" at (u,u'), in that it 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 very informative, although we encounter this phenomenon predominately for small samples and when true s_0 is very close to zero. To avoid this downward bias, we suggest to exclude problematic grid points with $\mathrm{Re}\left(H_T(u,u')\right) < 1$ 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 0 for a positive empirical covariance would correspond to setting $\mathrm{Re}\left(H_T(u,u')\right) = 1$ –, we simply drop the problematic points from the grid $\mathcal{U}$.

Remark 2. 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 that ε_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 ε_t. We thus do not pursue this idea any further. □

3.2 Large-Sample Properties of the Estimators

We now present the asymptotic properties of the various feasible estimators of s_0 proposed in Subsection 3.1.

3.2.1 Consistency and Asymptotic Normality

Assumption 2′. (i) Assumption 2 holds; and (ii) $1 \leq |\mathcal{U}| < \infty$.

Assumption 2′(ii) is assumed for easy implementation of our simple estimators.

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

Assumption 3. *The true unknown* s_0 *lies in the interior of* $\mathcal{S}$.

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 given in (74) in Appendix B. And D_0 is given in (26).

Theorem 4. *Suppose that Assumptions 1, 2′, 3 hold. Then:*

(i) $\sqrt{T}\left(\widehat{s}_{ecf} - s_0\right) \to^d \mathcal{N}\left(0, Asyvar\left(\widehat{s}_{ecf}\right)\right)$, *with*

$$Asyvar\left(\widehat{s}_{ecf}\right) := \left(\nabla_s R(\mathcal{U}; s_0)^\intercal D_0 \nabla_s R(\mathcal{U}; s_0)\right)^{-2} \times \nabla_s R(\mathcal{U}; s_0)^\intercal D_0 \Sigma_0 D_0 \nabla_s R(\mathcal{U}; s_0);$$

(ii) $\sqrt{T}\left(\widehat{s}_{ecf,2} - s_0\right) \to^d \mathcal{N}\left(0, Asyvar\left(\widehat{s}_{ecf,2}\right)\right)$, *with*

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

where Σ_0 *is a positive definite* $|\mathcal{U}| \times |\mathcal{U}|$ *matrix defined in Appendix C.*

3.2.2 The "Optimally" Weighted Estimator

For any positive semi-definite weight matrix $|\mathcal{U}| \times |\mathcal{U}|$ matrix D, and its consistent estimate $\widehat{D}_T$, we define an estimator $\widehat{s}_{ecf,\widehat{D}_T}$ as in (28).

Assumption 4. *(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 5. *Let 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*

$$Asyvar\left(\widehat{s}_{ecf,\widehat{D}_T}\right) := \left(\nabla_s R(\mathcal{U}; s_0)^\intercal D \nabla_s R(\mathcal{U}; s_0)\right)^{-2} \times \nabla_s R(\mathcal{U}; s_0)^\intercal D \Sigma_0 D \nabla_s R(\mathcal{U}; s_0). \tag{31}$$

(ii) Based on (31)*, the optimally weighed estimator of* s_0 *is given by*

$$\widehat{s}^*_{ecf} := \widehat{s}_{ecf,\widehat{\Sigma}_0^{-1}} = \arg\min_{s\in\mathcal{S}} Q_{\widehat{\Sigma}_0^{-1},T}\left(s, \mathcal{U}\right), \tag{32}$$

which satisfies $\sqrt{T}\left(\widehat{s}^*_{ecf} - s_0\right) \to^d \mathcal{N}\left(0, Asyvar\left(\widehat{s}^*_{ecf}\right)\right)$, *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, $Asyvar\left(\widehat{s}_{ecf}\right)$, $Asyvar\left(\widehat{s}_{ecf,2}\right)$, $Asyvar\left(\widehat{s}_{ecf,\widehat{D}_T}\right)$ 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, where $\widehat{s}$ 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 Appendix C.

Remark 3. When $|\mathcal{U}| = 1$, i.e., the grid $\mathcal{U}$ consists of a single point (u, u) with $0 < u < \overline{u}$, our estimation procedure has a closed-form solution that corresponds to Equation (10), i.e.,

$$\widehat{s}_{diag}(u) := \frac{2}{u}\arccos\left(\sqrt{|H_T(u,u)|^{-1}}\right). \tag{33}$$

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 estimator in (33) over various values of u could lead to efficiency gains. We leave this open for further research. □

Remark 4. One could drop Assumption 2′(ii) 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, our model is not first-order Markov and hence the semiparametric efficiency bound for s_0 is unknown. We leave it to future research for semiparametric efficient estimation of s_0.□

4 Simulation Studies and Empirical Application

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

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

We compare the finite sample performance 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 Equation (32); 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 $s_0 \in \{0.02, 0.2\}$ and $T = 250$ (this corresponds to roughly a year of daily closing prices). 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 four 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 re-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 $\overline{s} = 0.05$ (for $s_0 = 0.02$) and $\overline{s} = 0.5$ (for $s_0 = 0.2$), along with 500 equally spaced points in $[0, \overline{s}]$ 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 Roll's estimator we use two versions: $\widehat{s}_{Roll,1}$ denotes 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 Roll's estimator with Harris (1990) correction (i.e., use the absolute value of the empirical covariance).

- For 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 Hasbrouck's estimator where we set the estimate to zero in case the procedure does not converge; $\widehat{s}_{Has.,n^*=\cdot}$ denotes 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 light 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 for those two methods, whereas we expect our estimator to be more robust. The setup with (re-centered) log-normal innovations presents a regime with asymmetry, in which we expect Hasbrouck's estimator to be at a disadvantage. Indeed, these predictions 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 light-tailed regime, while both Roll's and 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; in particular, Hasbrouck's estimator is built on the assumption of normally distributed price innovations, which corresponds to the truth in this regime. However, Hasbrouck's estimator is sensitive and may be difficult to converge when the true unknown spread s_0 is large relative to the

variance of the latent price innovation.[6]

- 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 where our estimators beat those estimators by at least an order of magnitude in terms of RMSE. Although this case might be extreme, our empirical results in Section 4.2.1 suggest that for periods of heavy market turbulence this is not an unrealistic assumption. This makes the robustness of our estimator a relevant feature.

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

[6]For example, Hasbrouck's estimator only converges in about 60 % out of $n = 5,000$ simulation runs when $s_0 = 0.2$ and $\varepsilon_t \sim 0.02 \times \mathrm{N}(0, 1)$, which is consistent with its behavior 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}	$q_{97.5}$
$\varepsilon_t \sim 0.02\mathrm{N}(0,1)$	$\widehat{s}_{ecf}$	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
$\varepsilon_t \sim 0.02\mathrm{LN}(0,1.25)$	$\widehat{s}_{ecf}$	**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**
$\varepsilon_t \sim 0.02\mathrm{LN}(0,2)$	$\widehat{s}_{ecf}$	**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.)

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

In this section we illustrate the usefulness of our estimator with an application 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 underlying, 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 pm to 5 pm ET (Eastern Time), with a 15 min trading halt period Monday-Friday from 4:15 pm to 4:30 pm, and a maintenance period Monday-Thursday from 5 pm to 6 pm.[7]

In our application we look at the trading data for May 6, 2010.[8] During this day, financial markets in the U.S. experienced one of the most volatile periods on record, with major stock indices collapsing and rebounded within a short time frame of less than an hour.[9] 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 trading on May 6 happened around the time of the Flash Crash. For comparison purposes, 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%."

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

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

[8] Specifically, we look at all trades from May 5, 6 pm to May 6, 4:15 pm ET.

[9] 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).

Crash. We focus on the period from 2:32 pm to 3:08 pm 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., 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 involves consecutive trades at identical prices and price bounces in discrete (tick-size) steps. This makes the price innovations a discrete process, whereas 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 the estimation we use a rolling-window approach, where we estimate the spread for 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, for 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 its magnitude, the results a restated in basis points, BPS; 1 BPS = 1/100 %). The results are presented in Figure 4 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 between 2:45 pm to 2:49 pm 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 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 environment, suggests 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 outside the window of extreme turbulence both methods produce nearly identical results.

- 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. 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.

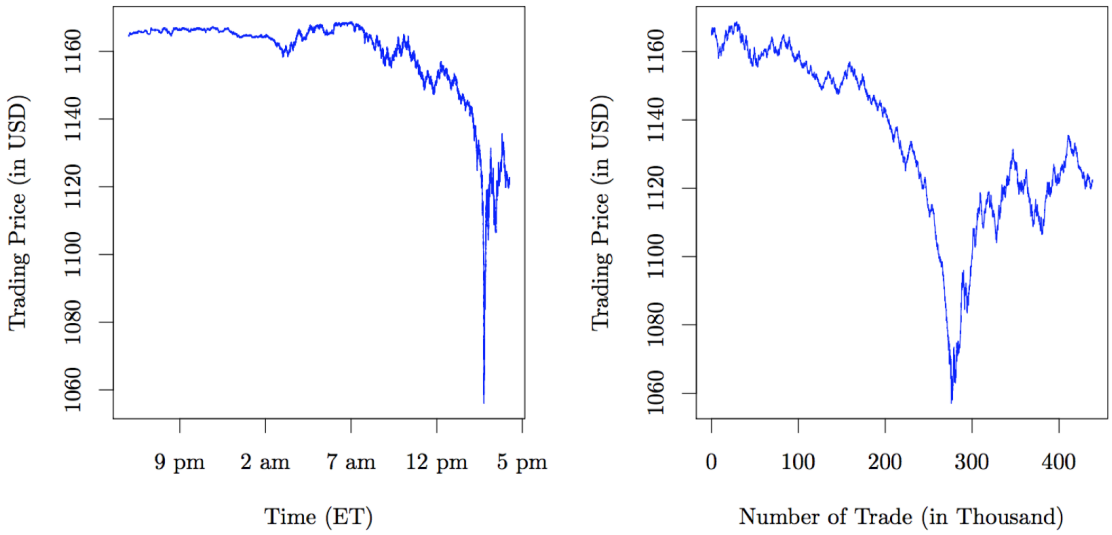

Figure 1: Transaction prices for E-Mini S&P futures (with maturity in June 2010) from May 5, 2010, 6 pm to May 6, 2010, 4:15 pm ET. Left: The last trading price for each second; Right: The sequence of all transaction prices throughout the day.

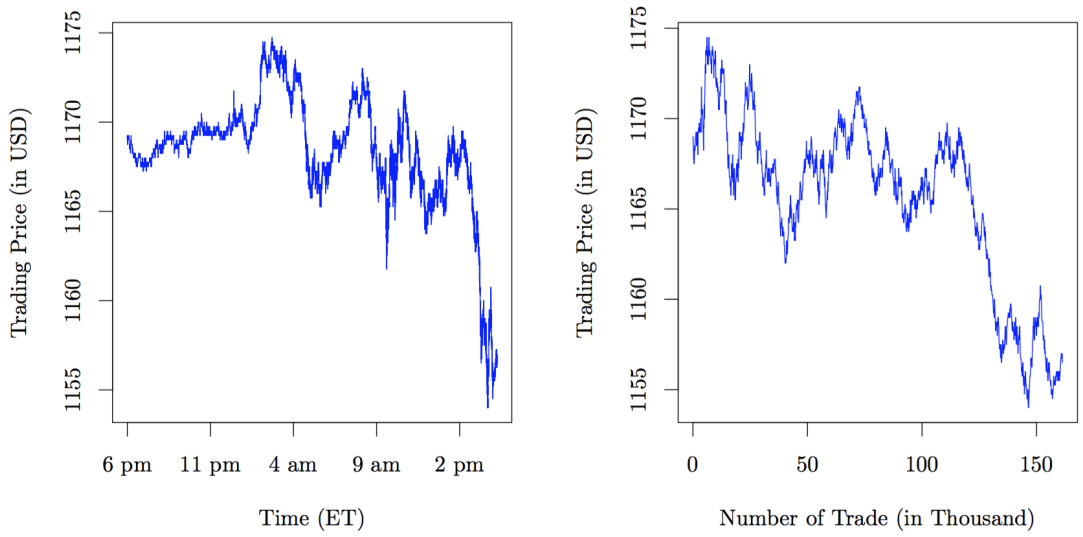

Figure 2: Transaction prices for E-Mini S&P futures (with maturity in June 2010) from May 12, 2010, 6 pm to May 13, 2010, 4:15 pm ET. Left: The last trading price for each second; Right: The sequence of all transaction prices throughout the day.

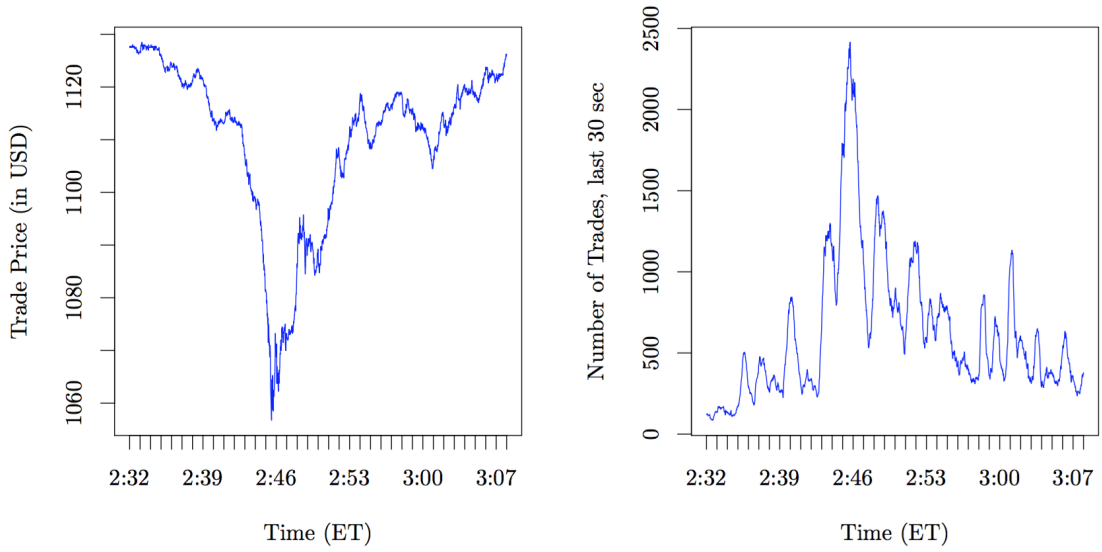

Figure 3: Transaction prices (left) and the number of trades in the last 30 seconds (right) for the period of the Flash Crash.

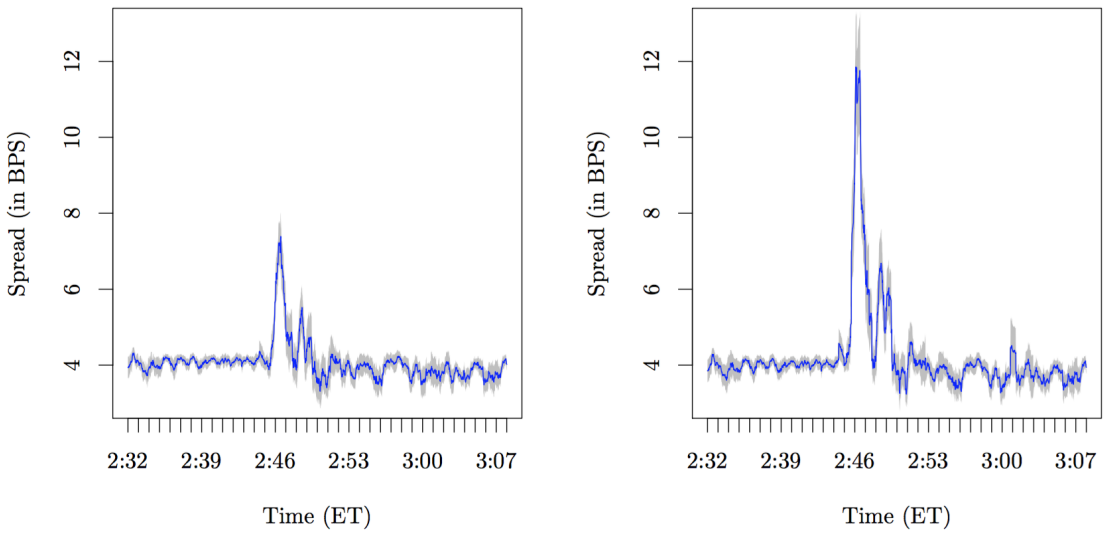

Figure 4: Spread estimates $\widehat{s}_{ecf}$ (left) and $\widehat{s}_{Roll,1}$ (right) for the period of the Flash Crash, with approximate 95 % confidence bands (gray area).

4.2.1 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 obtain estimates of the c.f. of the innovation process from (12) by using our spread estimator:

$$\widehat{\varphi}_\varepsilon(u) := \frac{\varphi_{T,1}(u)}{\left[\cos\left(u\frac{\widehat{s}_{ecf}}{2}\right)\right]^2}.$$

The properties of this estimator follow directly from our analysis of $\widehat{s}_{ecf}$ and from the properties of the sample characteristic function of the observed transaction prices. For an illustration, we incorporate this estimator into our empirical analysis of the E-mini futures data. We estimate the c.f. φ_ε for three different points in time: before, at, and after the spike in the estimated spread (see Figure 4); specifically, we choose the times 2:36 pm, 2:46 pm, and 2:56 pm 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 5:

- For 2:36 pm, we obtain an estimate that resembles the c.f. of a point mass at zero (i.e., a horizontal line), which is intuitive: the data shows 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 comes from randomly arriving buy and sell orders.

- However, during the turbulent period, when the spread peaks at around 2:46 pm, we obtain a significantly different behavior of the price innovations: the estimate declines in a nearly linear fashion, which corresponds to the c.f. of a heavy-tailed 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 innovations ε_t. In addition, this estimate supports our conjecture in Subsection 4.1 about the higher accuracy of our estimate $\widehat{s}_{ecf}$ during the turbulent period, compared to Roll's estimate $\widehat{s}_{Roll,1}$, which, based on simulation evidence, performs less well under heavy-tailed innovations.

- After the peak turbulence, at 2:56 pm, the estimate of the c.f. reflects a lighter-tailed regime again, close to the estimate that we obtain for 2:36 pm.

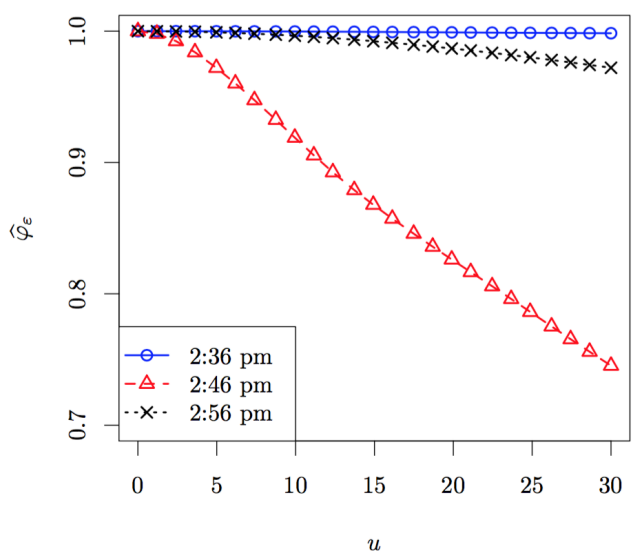

Figure 5: Estimates of the c.f. φ_ε of the fundamental price innovations ε_t during the Flash Crash of May 6, 2010. The plots/times refer to estimates before, at, and after the spike in the estimated spread, as displayed in Figure 4.

4.2.2 Detecting Order Flow Imbalances

The algebra in Section 5.1 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; on the other hand, under order flow imbalance (i.e., $\Pr(I_t = 1) \neq \Pr(I_t = -1)$), the quantity $H(u, u')$ is complex-valued when $u \neq u'$ with small $u' \neq 0$. This yields a way to detect order flow imbalances by measuring the imaginary part of the empirical quantity $H_T(u, u')$. In Figure 6, we plot the evolution of the two quantities

$$h_{max} := \max_{(u,u') \in \mathcal{U}} \left(\left| Im(H_T(u, u')) \right| \right) \qquad \text{and} \qquad h_{mean} := \frac{1}{|\mathcal{U}|} \sum_{(u,u') \in \mathcal{U}} \left(\left| Im(H_T(u, u')) \right| \right) \tag{34}$$

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 increases we detect), which indicates that not only the liquidity cost (as measured by the bid-ask spread) increased sharply, but that in addition the order flow became highly imbalanced during this period. This is in line with the economic intuition of a panic sale interpretation of the crash.

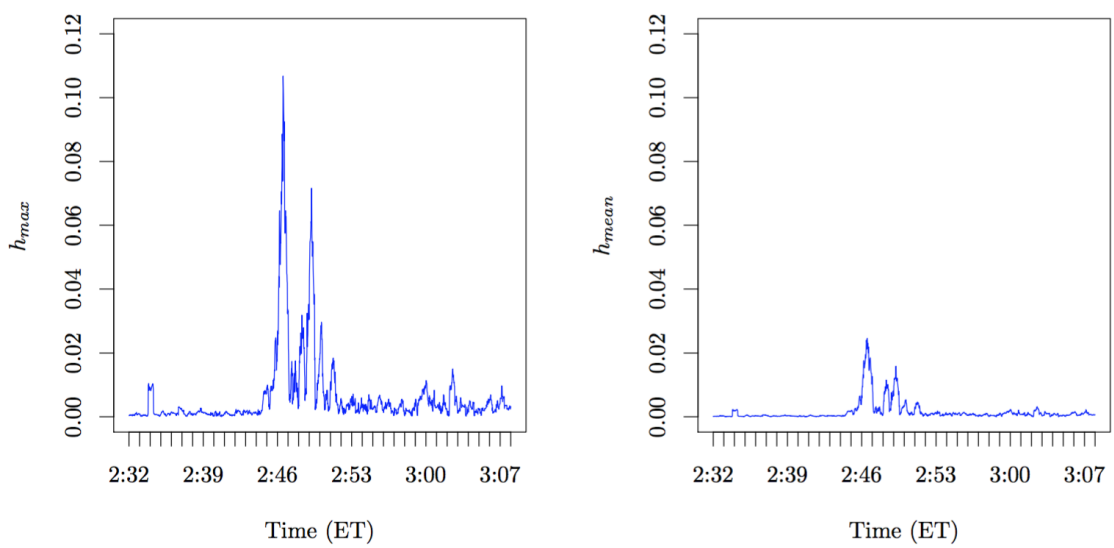

Figure 6: Indications of order flow imbalances during the Flash Crash of May 6, 2010. The definitions of the quantities h_{max} and h_{mean} are given in (34).

5 Extensions

This section presents identification results for four extended models that relax parts of Assumption 1 imposed in Sections 2 and 3. The purpose is to show how we may accommodate more general features in the basic Roll type model (1), and potentially how to estimate them from transaction data $\{p_t\}_{t=1}^T$ alone, without further observed information.

5.1 Unbalanced Order Flow

Assumption 5. *(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., $q_0 \neq 1/2$). Under Assumption 5, we obtain the following relations (similar to Equations (5), (6) and (7) in Section 2): for all $(u, u') \in \mathbb{R}^2$,

$$\begin{aligned}\varphi_{\Delta p,2}(u,u') &= \varphi_\varepsilon(u)\varphi_\varepsilon(u')\mathbb{E}\left(\exp\left(iu\frac{s_0}{2}I_t\right)\right)\mathbb{E}\left(\exp\left(i(u'-u)\frac{s_0}{2}I_{t-1}\right)\right)\mathbb{E}\left(\exp\left(-iu'\frac{s_0}{2}I_{t-2}\right)\right)\\ &= \varphi_\varepsilon(u)\varphi_\varepsilon(u')\left[\cos\left(u\frac{s_0}{2}\right)+(2q_0-1)i\sin\left(u\frac{s_0}{2}\right)\right]\left[\cos\left(u'\frac{s_0}{2}\right)-(2q_0-1)i\sin\left(u'\frac{s_0}{2}\right)\right]\\ &\times\left[\cos\left((u'-u)\frac{s_0}{2}\right)+(2q_0-1)i\sin\left((u'-u)\frac{s_0}{2}\right)\right],\end{aligned}$$

$$\begin{aligned}\varphi_{\Delta p,1}(u) &= \varphi_{\Delta p,2}(u,0) = \varphi_\varepsilon(u)\left[\cos^2\left(u\frac{s_0}{2}\right)+(2q_0-1)^2\sin^2\left(u\frac{s_0}{2}\right)\right],\\ \varphi_{\Delta p,2}(u,u) &= (\varphi_\varepsilon(u))^2\left[\cos^2\left(u\frac{s_0}{2}\right)+(2q_0-1)^2\sin^2\left(u\frac{s_0}{2}\right)\right].\end{aligned}$$

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

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

which is complex-valued unless either $u = u'$ or $(2q-1)\sin\left(u'\frac{s}{2}\right) = 0$. In particular, $R(u,u';s,1/2) = R(u,u';s)$ defined in Section 2. Similar to the identification Equation (14) for the basic Roll type model in Section 2, we have:

$$H(u,u') = R(u,u';s_0,q_0) \quad \textit{for all } (u,u') \in \overline{\mathcal{V}}^2, \tag{35}$$

and $H(u,u')$ is complex-valued unless either $u=u'$ or $(2q_0-1)\sin\left(u'\frac{s_0}{2}\right)=0$. Therefore for all $(\tilde{u},\tilde{u})\in\overline{\mathcal{V}}^2$ with $\tilde{u}\neq 0$, Equation (35) yields the relations

$$\begin{aligned} H(\tilde{u},\tilde{u}) &= \frac{1}{\cos^2\left(\tilde{u}\frac{s_0}{2}\right)+(2q_0-1)^2\sin^2\left(\tilde{u}\frac{s_0}{2}\right)}, \\ \iff \quad \cos^2\left(\tilde{u}\frac{s_0}{2}\right) &= \frac{1/H(\tilde{u},\tilde{u})-(2q_0-1)^2}{1-(2q_0-1)^2}, \end{aligned} \tag{36}$$

where $H(\tilde{u},\tilde{u})$ is real-valued with $H(\tilde{u},\tilde{u})>1$. Once $(2q_0-1)^2$ is identified or estimated, Equation (36) can be used to identify or estimate s_0 (as in Section 2). For $(\tilde{u},-\tilde{u})\in\overline{\mathcal{V}}^2$ with $\tilde{u}\neq 0$, Equation (35) implies

$$\frac{H(\tilde{u},-\tilde{u})}{[H(\tilde{u},\tilde{u})]^2}=\left[\cos\left(\tilde{u}\frac{s_0}{2}\right)+(2q_0-1)i\sin\left(\tilde{u}\frac{s_0}{2}\right)\right]^2\left[\cos\left(\tilde{u}s_0\right)-(2q_0-1)i\sin\left(\tilde{u}s_0\right)\right].$$

$$\begin{aligned} \mathrm{Re}\left(\frac{H(\tilde{u},-\tilde{u})}{[H(\tilde{u},\tilde{u})]^2}\right) &= (2q_0-1)^2+\left[(2q_0-1)^2-1\right]\cos^2\left(\tilde{u}\frac{s_0}{2}\right)\left[1-2\cos^2\left(\tilde{u}\frac{s_0}{2}\right)\right] \\ &= 2(2q_0-1)^2-H(\tilde{u},\tilde{u})^{-1}+2\frac{\left[H(\tilde{u},\tilde{u})^{-1}-(2q_0-1)^2\right]^2}{1-(2q_0-1)^2}, \end{aligned}$$

where the last equality uses the relation implied by Equation (36). Therefore,

$$(2q_0-1)^2=\frac{\mathrm{Re}\left(\frac{H(\tilde{u},-\tilde{u})}{[H(\tilde{u},\tilde{u})]^2}\right)+H(\tilde{u},\tilde{u})^{-1}-2H(\tilde{u},\tilde{u})^{-2}}{2+\mathrm{Re}\left(\frac{H(\tilde{u},-\tilde{u})}{[H(\tilde{u},\tilde{u})]^2}\right)-3H(\tilde{u},\tilde{u})^{-1}} \tag{37}$$

which can be used to identify and estimate $(2q_0-1)^2$.

$$\begin{aligned} \mathrm{Im}\left(\frac{H(\tilde{u},-\tilde{u})}{[H(\tilde{u},\tilde{u})]^2}\right) &= \left[(2q_0-1)^2-1\right](2q_0-1)\sin^2\left(\tilde{u}\frac{s_0}{2}\right)\sin\left(\tilde{u}s_0\right) \\ &= 2(1-2q_0)(1-H(\tilde{u},\tilde{u})^{-1})\sqrt{\frac{1/H(\tilde{u},\tilde{u})-(2q_0-1)^2}{1-(2q_0-1)^2}}\sqrt{\frac{1-1/H(\tilde{u},\tilde{u})}{1-(2q_0-1)^2}}, \end{aligned} \tag{38}$$

which can be used to identify the sign of $2q_0-1$ for a small $\tilde{u}\neq 0$. These arguments lead to the following theorem.

Assumption 6. *(i) Assumption 2(i) holds; (ii) either (a) $\mathcal{U}=\overline{\mathcal{U}}$; or (b) $\mathcal{U}\subset\overline{\mathcal{U}}$, and $\exists(\tilde{u},\tilde{u}),(\tilde{u},-\tilde{u})\in\mathcal{U}$ such that $0<\tilde{u}<\overline{u}$, where $\overline{u}$ denotes the first positive zero of $u\mapsto\min_{s\in\mathcal{S}}\cos\left(u\frac{s}{2}\right)$.*

Theorem 6. *(1) Let Assumption 5 hold. Then: the c.f. $\varphi_\varepsilon(\cdot)$ is identified as (9) on $\overline{\mathcal{V}}$, and (s_0,q_0) is identified by Equations (36), (37) and (38) with a small positive $\tilde{u}\in\overline{\mathcal{V}}$.*
(2) Let Assumptions 5 and 6 hold. Then: (s_0,q_0) is identified as the unique solution to the minimum distance criterion function based on Equation (35) evaluated on $\mathcal{U}$.

In Theorem 6 part(2), the minimum distance criterion function can be constructed similar to Equation (27). Then the consistency and the asymptotic normality are readily established similar to Theorems 3, 4 and 5. 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 implemented in the empirical application section 4.2.

5.2 Model when $\{I_t\}$ has general discrete support

We now consider a generalization of the Roll model by relaxing Assumption 1(iii) on the support of the latent trade direction indicators.

Assumption 7. *(i) Assumption 1(i)(ii) holds; and (ii) $\{I_t\}$ may take values in $\{-k_1, \ldots, 0, \ldots, +k_2\}$, and $\Pr(I_t = -k_1) > 0$, $\Pr(I_t = +k_2) > 0$.*

Here, k_1 and k_2 are positive integers, measuring the strength of the order flow. Assumption 7(ii) allows the case where $\Pr(I_t = 0) = 0$ or $\Pr(I_t = 0) > 0$. It also allows for asymmetric support in the sense that $k_1 \neq k_2$. Denote the unknown marginal probabilities of $\{I_t\}$ as $\pi_0 = [\vec{\pi_l}]$, where $\pi_l = \Pr(I_t = l) \geq 0$, for $l = -k_1, \ldots, 0, \ldots, +k_2$ and $\sum_l \pi_l = 1$. Let $\varphi_I(u) := \mathbb{E}\left(\exp\left(iuI_t\right)\right)$ denote the c.f. of I_t, which is analytic and is uniquely determined by the unknown π_0. By the inversion theorem, the unknown π_0 is identified as long as its c.f. $\varphi_I(\cdot)$ is identified. Under Assumption 7(i) (i.e., Assumption 1(i)(ii)), we obtain: for all $(u, u') \in \mathbb{R}^2$,

$$\varphi_{\Delta p,2}(u, u') = \varphi_\varepsilon(u)\varphi_\varepsilon(u')\varphi_I\left(u\frac{s_0}{2}\right)\varphi_I\left((u'-u)\frac{s_0}{2}\right)\varphi_I\left(-u'\frac{s_0}{2}\right), \tag{39}$$

$$\varphi_{\Delta p,1}(u) = \varphi_\varepsilon(u)\varphi_I\left(u\frac{s_0}{2}\right)\varphi_I\left(-u\frac{s_0}{2}\right), \tag{40}$$

$$\varphi_{\Delta p,2}(u, u) = \varphi_\varepsilon(u)\varphi_\varepsilon(u)\varphi_I\left(u\frac{s_0}{2}\right)\varphi_I\left(-u\frac{s_0}{2}\right). \tag{41}$$

By Equations (41) and (40), the c.f. $\varphi_\varepsilon(\cdot)$ is identified as (9) on $\overline{\mathcal{V}}$. Denote

$$R(u, u'; s_0, \pi_0) := \frac{\mathbb{E}\left(\exp\left[i\frac{s_0}{2}\left(u'-u\right)I_{t-1}\right]\right)}{\mathbb{E}\left(\exp\left[iu'\frac{s_0}{2}I_{t-1}\right]\right)\mathbb{E}\left(\exp\left[-iu\frac{s_0}{2}I_{t-1}\right]\right)}.$$

Then Equations (39) and (40) imply the following relation:

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

Under Assumption7(ii), we prove in Appendix that Equation (42) identifies both s_0 and $\varphi_I(\cdot)$.

Theorem 7. *(1) Let Assumption 7(i) hold. Then the c.f. $\varphi_\varepsilon(\cdot)$ is identified as* (9) *on $\overline{\mathcal{V}}$.*
(2) Let Assumption 7(i)(ii) hold. Then: s_0 and the c.f. $\varphi_I(\cdot)$ are identified.

We can jointly estimate s_0 and π_0 by essentially the same minimum distance strategy as in Section 3 based on an empirical version of the identification equation (42). Recently Zhang and Hodges (2012) consider a model where our Assumption 7(ii) is replaced by $\{I_t\}$ having support in $\{-\lambda, -1, 1, \lambda\}$. They do not study the identification issue but directly apply Bayesian Gibbs method for estimation under the additional assumption of $\varepsilon_t \sim$ i.i.d. $\mathrm{N}(0, \sigma_\varepsilon)$.

5.3 General model when $\{I_t\}$ is a Stationary Markov Chain of Order 1

$\{I_t\}$ could also be a stationary first-order Markov Chain with

$$\Pr(I_t = j | I_{t-1} = m) = q_{mj}, \quad for\ m = -k, \ldots, 0, \ldots, +k,\ and\ j = -k, \ldots, 0, \ldots, +k.$$

The probabilistic property of $\{I_t\}$ is determined by the unknown transition matrix $\mathcal{Q}_0 = [q_{mj}]$. Denote the associated stationary marginal probabilities of $\{I_t\}$ as $\pi_0 = [\vec{\pi}_l]$, where $\pi_l = \Pr(I_t = l)$, for $l = -k, \ldots, 0, \ldots, +k$ and $\sum_l \pi_l = 1$.

Assumption 8. *(i) Assumption 1(i) holds; (ii) $\{I_t\}$ is strictly stationary first-order Markov, irreducible and aperiodic; and (iii) $\{I_t\}$ takes values in $\{-k, \ldots, 0, \ldots, +k\}$, and $\Pr(I_t = l) > 0$, for $l = -k, \ldots, 0, \ldots, +k$.*

Since $\{I_t\}$ is a strictly stationary, finite-state Markov chain, by Theorem 3.1 of Bradley (2005), $\{I_t\}$ being irreducible and aperiodic is equivalent to its being $\psi-$mixing or strongly mixing. And under such condition, the mixing rates are (at least) exponentially fast. Assumption 8 (ii) is assumed, because $\{\Delta p_t\}$ is observed to be stationary and display short memory in real data. Under Assumption 8, we obtain for all $(u, u') \in \mathbb{R}^2$,

$$\varphi_{\Delta p,1}(u) = \varphi_\varepsilon(u)\mathbb{E}\left(\exp\left[iu\frac{s_0}{2}\left(I_t - I_{t-1}\right)\right]\right),$$
$$\varphi_{\Delta p,2}(u, u') = \varphi_\varepsilon(u)\varphi_\varepsilon(u')\mathbb{E}\left(\exp\left[iu\frac{s_0}{2}\left(I_t - I_{t-1}\right)\right]\exp\left[iu'\frac{s_0}{2}\left(I_{t-1} - I_{t-2}\right)\right]\right).$$

Suppose $\theta_0 = (s_0, \mathcal{Q}_0) \in \Theta \subset \mathbb{R}^d$. It follows that

$$H(u, u') = R(u, u'; s_0, \mathcal{Q}_0), \qquad for\ all\ (u, u') \in \overline{\mathcal{V}}^2, \tag{43}$$

where

$$R(u,u';s_0,\mathcal{Q}_0) := \frac{\mathbb{E}\left(\exp\left[iu\frac{s_0}{2}\left(I_t - I_{t-1}\right)\right]\exp\left[iu'\frac{s_0}{2}\left(I_{t-1} - I_{t-2}\right)\right]\right)}{\mathbb{E}\left(\exp\left[iu\frac{s_0}{2}\left(I_t - I_{t-1}\right)\right]\right)\mathbb{E}\left(\exp\left[iu'\frac{s_0}{2}\left(I_{t-1} - I_{t-2}\right)\right]\right)}.$$

We can identify $\theta_0 = (s_0, \mathcal{Q}_0) \in \Theta \subset \mathbb{R}^d$ by considering lots of $(u, u') \in \overline{\mathcal{V}}^2$. We next establish identification. Let $\varphi_{\Delta I}(\cdot,\cdot)$ denote the true unknown joint c.f. of $(I_{t-1} - I_{t-2}, I_t - I_{t-1})$. In the following lemma, we first establish the identification result for the joint distribution of $(I_{t-1} - I_{t-2}, I_t - I_{t-1})$ and the spread, i.e.$(s_0, \varphi_{\Delta I}(\cdot,\cdot))$. Since $\{I_t\}$ takes values in $\{-k,\ldots,0,\ldots,+k\}$, the support of $(I_t - I_{t-1})$ is $\{-2k,\ldots,0,\ldots,+2k\}$ and the joint support of $(I_{t-1} - I_{t-2}, I_t - I_{t-1})$ is

$$\begin{bmatrix}
 & & & & (-2k,0) & \cdots & \cdots & \cdots & (-2k,2k) \\
 & & & (-2k+1,-1) & (-2k+1,0) & \cdots & \cdots & \cdots & (-2k+1,2k) \\
 & & (-2k+2,-2) & (-2k+2,-1) & (-2k+2,0) & \cdots & \cdots & \cdots & (-2k+2,2k) \\
 & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & (-1,-2k+1) & \cdots & \cdots & (-1,0) & \cdots & \cdots & (-1,2k-1) & (-1,2k) \\
(0,-2k) & (0,-2k+1) & \cdots & \cdots & (0,0) & \cdots & & \cdots(0,2k-1) & (0,2k) \\
(1,-2k) & (1,-2k+1) & \cdots & \cdots & (1,0) & \cdots & \cdots & (1,2k-1) & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \\
(2k-2,-2k) & \cdots & \cdots & \cdots & (2k-2,0) & (2k-2,1) & (2k-2,2) & & \\
(2k-1,-2k) & \cdots & \cdots & \cdots & (2k-1,0) & (2k-1,1) & & & \\
(2k,-2k) & \cdots & \cdots & \cdots & (2k,0) & & & &
\end{bmatrix}. \tag{44}$$

When one uses Equation (43) for estimation, the joint support information given in Equation (44) shall be used to improve efficiency. Denote the joint probability mass matrix of $(I_{t-1} - I_{t-2}, I_t - I_{t-1})$ as $\mathcal{Q}^0_{\Delta I}$, which is a $(4k+1)\times(4k+1)$ matrix. Denote the row vectors of $\mathcal{Q}_0$ as $Q_{j,\cdot} = [q_{j,-k},\cdots,q_{j,k}]$, for $j = -k,\ldots,0,\ldots,+k$. The summation of each component of $Q_{j,\cdot}$ equals to 1, according to the definition. The following equation shows the connection between $\mathcal{Q}^0_{\Delta I}$ and $\mathcal{Q}_0$, π_0 :

$$\mathcal{Q}^0_{\Delta I} = A_{\mathcal{Q}_0,\pi_0} \times B_{\mathcal{Q}_0}, \tag{45}$$

where $A_{\mathcal{Q}_0,\pi_0}$ is a $(4k+1)\times(2k+1)$ matrix

$$\begin{bmatrix}
\pi_k q_{k,-k} & 0 & 0 & 0 & \cdots & \cdots & 0 \\
\pi_{k-1} q_{k-1,-k} & \pi_k q_{k,-k+1} & 0 & 0 & \cdots & \cdots & 0 \\
\pi_{k-2} q_{k-2,-k} & \pi_{k-1} q_{k-1,-k+1} & \pi_k q_{k,-k+2} & 0 & \cdots & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\pi_{-k+2} q_{-k+2,-k} & \pi_{-k+3} q_{-k+3,-k+1} & \pi_{-k+4} q_{-k+4,-k+2} & \cdots & \pi_k q_{k,k-2} & 0 & 0 \\
\pi_{-k+1} q_{-k+1,-k} & \pi_{-k+2} q_{-k+2,-k+1} & \pi_{-k+3} q_{-k+3,-k+2} & \cdots & \pi_{k-1} q_{k-1,k-2} & \pi_k q_{k,k-1} & 0 \\
\pi_{-k} q_{-k,-k} & \pi_{-k+1} q_{-k+1,-k+1} & \pi_{-k+2} q_{-k+2,-k+2} & \cdots & \pi_{k-2} q_{k-2,k-2} & \pi_{k-1} q_{k-1,k-1} & \pi_k q_{k,k} \\
0 & \pi_{-k} q_{-k,-k+1} & \pi_{-k+1} q_{-k+1,-k+2} & \cdots & \pi_{k-3} q_{k-3,k-2} & \pi_{k-2} q_{k-2,k-1} & \pi_{k-1} q_{k-1,k} \\
0 & 0 & \pi_{-k} q_{-k,-k+2} & \cdots & \pi_{k-4} q_{k-4,k-2} & \pi_{k-3} q_{k-3,k-1} & \pi_{k-2} q_{k-2,k} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & 0 & \pi_{-k} q_{-k,k-2} & \pi_{-k+1} q_{-k+1,k-1} & \pi_{-k+2} q_{-k+2,k} \\
0 & \cdots & 0 & 0 & 0 & \pi_{-k} q_{-k,k-1} & \pi_{-k+1} q_{-k+1,k} \\
0 & \cdots & 0 & 0 & 0 & 0 & \pi_{-k} q_{-k,k}
\end{bmatrix},$$

and $B_{\mathcal{Q}_0}$ is $(2k+1)\times(4k+1)$ matrix

$$\begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 & 0 & 0 & Q_{-k,\cdot} \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 & 0 & Q_{-k+1,\cdot} & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 & Q_{-k+2,\cdot} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & Q_{-1,\cdot} & \cdots & 0 & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & Q_{0,\cdot} & 0 & \cdots & 0 & 0 \\ 0 & \cdots & \cdots & 0 & Q_{1,\cdot} & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & Q_{k-2,\cdot} & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & Q_{k-1,\cdot} & 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ Q_{k,\cdot} & 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \end{bmatrix}.$$

Thus the rank of $\mathcal{Q}^0_{\Delta I}$ is at most $2k+1$. Since it does not satisfy the non-singularity condition, Theorem 1 of Gassiat and Rousseau (2016) could not be applied to our case. Once $\varphi_{\Delta I}(\cdot,\cdot)$ or equivalently $\mathcal{Q}^0_{\Delta I}$ is identified and estimated, Equation (45) can be used to recover $\mathcal{Q}_0$ and π_0.

Assumption 9. *(i) $A_{\mathcal{Q}_0,\pi_0}$ is of full column rank; and (ii) $q_{-k,-k} > \frac{1}{2}$ and $q_{k,k} > \frac{1}{2}$.*

For example, if $\pi_k q_{k,j} > 0$, for $j = -k,\cdots,k$, or $\pi_{-k} q_{-k,j} > 0$, for $j = -k,\cdots,k$, then Assumption 9(i) is satisfied. Also, when $k=1$ (as in the basic Roll model), Assumption 9(ii) could be interpreted as a model of (time-varying) autocorrelation in the trade indicators: after a buy, the most likely thing is another buy, and analogously for a sell.

Lemma 1. *Suppose that Assumptions 8 and 9 hold. Then on $\overline{\mathcal{V}}^2$, $(s_0, \varphi_{\Delta I}(\cdot,\cdot))$ and the c.f. φ_ε are uniquely identified.*

In general, $(s_0, \varphi_{\Delta I}(\cdot,\cdot))$ cannot be identified, without information about the support.

Example 5.1. *$\{I_t\}$ could take values in $\{-2,-1,0,1,2\}$. The marginal distribution satisfies $\Pr(I_t = -1) = \Pr(I_t = 1) = 1/2$, and the transition matrix is $[1/3\ 2/3; 2/3\ 1/3]$. Define $W_t = 1/2\,[I_t - I_{t-1} + e_t]$, with $\{e_t\}$ being independent of $\{I_t\}$, and $\Pr(e_t = -2) = b, \Pr(e_t = 2) = 1-b$. It is easy to show the joint support of (W_{t-1}, W_t) is a subset of Equation (44) for $k=2$. Therefore, Equation (43) cannot distinguish $(s, \varphi_{\Delta I}(\cdot,\cdot))$ from $(2s, \varphi_W(\cdot,\cdot))$, where $\varphi_W(\cdot,\cdot)$ is the joint c.f.of (W_{t-1}, W_t). Simple calculations show $\Pr(W_{t-1} = -2, W_t = -1) = \Pr(W_{t-1} = -1, W_t = -2) = \frac{1}{9}b^2 > 0$, $\Pr(W_{t-1} = 1, W_t = 2) = \Pr(W_{t-1} = 2, W_t = 1) = \frac{1}{9}(1-b)^2 > 0$. If one has additional information that $\Pr(I_t = -2) = \Pr(I_t = 2) = 0$, then it is known that $(-2,-1)$, $(-1,-2)$, $(1,2)$, $(2,1)$, are not in Equation (44) for $k=1$. Thus one is able to distinguish $(s, \varphi_{\Delta I}(\cdot,\cdot))$ from $(2s, \varphi_W(\cdot,\cdot))$. More*

generally, let $W_t = c\,[I_t - I_{t-1} + e_t]$, where c is any constant and $\{e_t\}$ is independent of $\{I_t\}$. The joint support of (W_{t-1}, W_t) is not a subset of Equation (44) for $k = 1$.

We next establish the identification results for the joint distribution of (I_{t-1}, I_t).

Theorem 8. *Suppose that Assumptions 8 and 9 hold. Furthermore, $q_{k,-j} > 0$, for $j = 1, \cdots, k$ and $q_{-k,j} > 0$, for $j = 0, 1, \cdots, k$. Then s_0 and the joint distribution of (I_{t-1}, I_t) are uniquely identified.*

Lemma 1 shows the identification result for $\mathcal{Q}^0_{\Delta I}$. Since $\mathcal{Q}^0_{\Delta I} = A_{\mathcal{Q}_0,\pi_0} \times B_{\mathcal{Q}_0}$, we show in the proof of Theorem 8 that $B_{\mathcal{Q}_0}$ or equivalently the joint distribution of (I_{t-1}, I_t) can be solved back under some conditions on $A_{\mathcal{Q}_0,\pi_0}$. Theorem 8 only gives one such sufficient condition.

5.4 Adverse Selection

In all the above extensions we have assumed that the price dynamics follows Equation (1). We now relax this condition and suppose that

$$\Delta p_t = \varepsilon_t + \alpha_0 I_t - \beta_0 I_{t-1}, \tag{46}$$

where the other parts of Assumption 1 are kept. This equation arises from considering the presence of an adverse selection component in the spread, see Equation (5.4) in Foucault et al. (2013). In this case, $\beta_0 = s_0/2$ and $\alpha_0 = s_0/2 + \delta$, where $\delta = \alpha_0 - \beta_0 \neq 0$ measures the contribution of adverse selection. Rewriting (46) in the form of our previous price dynamics in (1), 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 $\mathrm{Cov}\,(\tilde{\varepsilon}_t, I_t) = \delta \neq 0$, so that our estimator (and the Roll and Hasbrouck estimators) would be biased. Under only autocovariance restrictions and without trade direction data, $(\alpha_0, \beta_0, \sigma^2_\varepsilon)$ cannot be jointly identified (even under Hasbrouck (2004)'s assumption of $\varepsilon_t \sim$ i.i.d. $\mathrm{N}(0, \sigma_\varepsilon)$). We now show how to obtain identification under Hasbrouck (2004)'s stronger independence assumption that $\{\varepsilon_t\}$ is independent of $\{I_t\}$.

Assumption 10. *(i) Data $\{p_t\}_{t=1}^T$ is generated from Equation (46) with $\beta_0 > 0$, where $\{\varepsilon_t\}$ is i.i.d. and independent of $\{I_t\}$; (ii) Assumption 1(ii)(iii) holds.*

This assumption implies that for all $(u, u') \in \mathbb{R}^2$,

$$\begin{aligned}\varphi_{\Delta p,2}(u,u') &= \varphi_\varepsilon(u)\varphi_\varepsilon(u')\mathbb{E}\left(\exp\left(iu\alpha_0 I_t\right)\right)\mathbb{E}\left(\exp\left(i(u'\alpha_0 - u\beta_0)I_{t-1}\right)\right)\mathbb{E}\left(\exp\left(-iu'\beta_0 I_{t-2}\right)\right)\\ &= \varphi_\varepsilon(u)\varphi_\varepsilon(u')\cos\left(u\alpha_0\right)\cos\left(u'\alpha_0 - u\beta_0\right)\cos\left(u'\beta_0\right), \qquad (47)\\ \varphi_{\Delta p,1}(u) &= \varphi_{\Delta p,2}(u,0) = \varphi_\varepsilon(u)\cos\left(u\alpha_0\right)\cos\left(u\beta_0\right). \qquad (48)\end{aligned}$$

Denote

$$\overline{\mathcal{U}}_{as} := \left\{ (u, u') \in \overline{\mathcal{V}}^2 : \min_{(\alpha,\beta)\in\mathcal{S}^2} |\cos(u\beta)\cos(u'\alpha)| > 0 \right\}, \tag{49}$$

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

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

Equations (47) and (48) now imply that

$$H(u, u') = R(u, u'; \alpha_0, \beta_0) \quad for \ (u, u') \in \overline{\mathcal{V}}^2, \tag{50}$$

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 $(0, 0)$, for a small positive $\tilde{u} \in \overline{\mathcal{V}}$, we have $(\tilde{u}, \tilde{u}), (\tilde{u}, 2\tilde{u}), (2\tilde{u}, \tilde{u}) \in \mathcal{V}$, and Equation (50) yields

$$\sin^2(\tilde{u}\alpha_0) = \frac{2H(\tilde{u}, \tilde{u}) - H(\tilde{u}, 2\tilde{u}) - 1}{2H(\tilde{u}, \tilde{u}) - 2H(\tilde{u}, 2\tilde{u})}, \tag{51}$$

$$\sin^2(\tilde{u}\beta_0) = \frac{2H(\tilde{u}, \tilde{u}) - H(2\tilde{u}, \tilde{u}) - 1}{2H(\tilde{u}, \tilde{u}) - 2H(2\tilde{u}, \tilde{u})}. \tag{52}$$

Since $0 < \tilde{u} < \frac{\overline{u}}{2}$, $s \mapsto \sin^2(\tilde{u}s)$ is strictly increasing in $s \in \mathcal{S}$. This implies that (51) and (52) hold only at $\alpha = \alpha_0$ and $\beta = \beta_0$. Consequently,

$$\alpha_0 = \tilde{u}^{-1} \arcsin\left(\sqrt{\frac{2H(\tilde{u}, \tilde{u}) - H(\tilde{u}, 2\tilde{u}) - 1}{2H(\tilde{u}, \tilde{u}) - 2H(\tilde{u}, 2\tilde{u})}}\right), \ \beta_0 = \tilde{u}^{-1} \arcsin\left(\sqrt{\frac{2H(\tilde{u}, \tilde{u}) - H(2\tilde{u}, \tilde{u}) - 1}{2H(\tilde{u}, \tilde{u}) - 2H(2\tilde{u}, \tilde{u})}}\right).$$

Assumption 11. *(i) $(\alpha_0, \beta_0) \in \mathcal{S}^2$; (ii) either (a) $\mathcal{U} = \overline{\mathcal{U}}_{as}$; or (b) $\mathcal{U} \subset \overline{\mathcal{U}}_{as}$ and $\exists(\tilde{u}, \tilde{u}), (\tilde{u}, 2\tilde{u}), (2\tilde{u}, \tilde{u}) \in \mathcal{U}$, such that $0 < \tilde{u} < \frac{\overline{u}}{2}$, where $\overline{u}$ denotes the first positive zero of $u \mapsto \min_{s\in\mathcal{S}} \cos(us)$.*

Theorem 9. *(1) Let Assumption 10 hold. Then: (α_0, β_0) is identified by Equations (51) and (52) for a small positive $\tilde{u} \in \overline{\mathcal{V}}$, and the c.f. φ_ε is identified on $\overline{\mathcal{V}}$ as $\varphi_\varepsilon(u) = \frac{\varphi_{\Delta p,1}(u)}{\cos(u\alpha_0)\cos(u\beta_0)}$.*
(2) Let Assumptions 10 and 11 hold. Then: (α_0, β_0) is identified as the unique solution to the minimum distance criterion function based on Equation (50) evaluated on $\mathcal{U}$.

From this we can identify (s_0, δ) jointly. In Theorem 9 part (2), the minimum distance criterion function can be constructed similar to Equation (27). The consistency and the asymptotic normality are readily established similar to Theorems 3, 4 and 5. We note that, with additional data such as trade volume, the independence between $\{\varepsilon_t\}$ and $\{I_t\}$ condition in Assumption 10 could be dropped. Furthermore, one could modify some recently development in nonclassical measurement error (see, e.g., Hu (2016), Carroll et al. (2006)) to obtain further extensions of the basic Roll model.

6 Conclusions

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

In our application to the E-mini futures contract on the S&P 500 during the Flash Crash, we find that during relatively tranquil times our estimator $\widehat{s}_{ecf}$ and the Roll estimator $\widehat{s}_{Roll,1}$ are very similar, but during the peak period of the Flash Crash, i.e., between 2:45 pm to 2:49 pm 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 turbulence in market prices during this period, along with the simulation evidence on the robustness of our estimator $\widehat{s}_{ecf}$ in a heavy-tailed environment, suggests that $\widehat{s}_{Roll}$ 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 outside the window of extreme turbulence both methods produce nearly identical results. We also found that order flow became badly unbalanced. Both of these findings corroborate the work presented in the SEC/CFTC report on the days events and subsequent academic work. We also find however that the fundamental innovation became much more heavy tailed during the critical period, so that perhaps explanations are not just due to market structure related issues.

We have emphasized in the theoretical treatment the plain Roll model, but we also showed how certain extensions such as unbalanced order flow, or serially dependent latent trade indicators, or adverse selection can be well accommodated in our framework. In fact, it may be possible to consider further extensions that allow several features all at once, and one could consider more efficient estimators as well. We leave these for future work.

Appendices

The Appendices consist of all the proofs and additional figures.

A. Proofs for Identification Results in Sections 2 and 5

Proof of Theorem 2 Both criterion functions (18) and (19) are nonnegative, with $J(s_0,\mathcal{U}) = Q(s_0,\mathcal{U}) = 0$, under Assumption 2(ii). For either case of Assumption 2(ii), $\exists(\tilde{u},\tilde{u}) \in \mathcal{U}$ with $\tilde{u} > 0$. For this grid point, the moment condition (15) yields the relation

$$\cos^2\left(\tilde{u}\frac{s_0}{2}\right) = \frac{\varphi^2_{\Delta p,1}(\tilde{u})}{\varphi_{\Delta p,2}(\tilde{u},\tilde{u})}. \tag{53}$$

By Assumption 2(ii), $\tilde{u}$ is smaller than the first positive zero of $u \mapsto \min_{s\in\mathcal{S}} \cos\left(u\frac{s}{2}\right)$, and hence $s \mapsto \cos^2\left(\tilde{u}\frac{s}{2}\right)$ is strictly decreasing in $s \in \mathcal{S}$. This implies that (53) holds only at $s = s_0$, which further implies that both criterion functions are uniquely minimized at $s = s_0$. This gives the identification result. **Proof of Theorem 7**

Let φ_I denote the c.f.of $\{I_t\}$ associated with the true unknown π_0, which is analytic on $\mathbb{R}$, since $\{I_t\}$ is discrete with support $\{-k_1,\ldots,0,\ldots,+k_2\}$. Equation (42) gives on $\overline{\mathcal{V}}^2$,

$$H(u,u') = \frac{\varphi_I\left(\frac{s_0}{2}(u'-u)\right)}{\varphi_I\left(\frac{s_0}{2}u'\right)\varphi_I\left(-\frac{s_0}{2}u\right)}. \tag{54}$$

If the pair $(\tilde{s} \in \mathcal{S}, \psi(\cdot))$ also satisfies Equation (42), where ψ denotes the c.f.of $\{I_t\}$ associated with another probability mass function π i.e.,

$$H(u,u') = \frac{\varphi_I\left(\frac{s_0}{2}(u'-u)\right)}{\varphi_I\left(\frac{s_0}{2}u'\right)\varphi_I\left(-\frac{s_0}{2}u\right)} = \frac{\psi\left(\frac{\tilde{s}}{2}(u'-u)\right)}{\psi\left(\frac{\tilde{s}}{2}u'\right)\psi\left(-\frac{\tilde{s}}{2}u\right)}. \tag{55}$$

Below we shall prove that $\varphi_I(\frac{s_0}{2}u) = \exp(ifu)\psi\left(\frac{\tilde{s}}{2}u\right)$, where $f \in \mathbb{R}$ is a constant, with no information about the support of $\{I_t\}$. This result is intuitive. Since we only have observations for $\frac{s_0}{2}(I_t - I_{t-1})$, we could not differentiate between I_t and $I_t + f$, for a constant f, or between (I_t, s_0) and $(I_t \cdot \frac{s_0}{\tilde{s}}, \tilde{s})$, for a positive constant $\tilde{s}$, without additional information about the support. Assumption 7 (ii) excludes the possibility of a change of the location or the scale, then $\theta_0 = (s_0, \pi_0^\intercal)^\intercal$ can be uniquely identified from Equation (42). Denote $h(u) = \psi\left(\frac{\tilde{s}}{s_0}u\right)$, and $u_1 = -\frac{s_0}{2}u$, $u_2 = \frac{s_0}{2}u'$. Note that $\varphi_I(\cdot)$, $\psi(\cdot)$, $h(\cdot)$ are analytic on $\mathbb{R}$ and equal to 1 at 0. There exists a small neighbourhood $\mathcal{M}$ of $(0,0) \subset \overline{\mathcal{V}}^2$, such that $\varphi_I(u_1)$, $\varphi_I(u_2)$, $\varphi_I(u_1+u_2)$, $h(u_1)$, $h(u_2)$ and $h(u_1+u_2)$ are all

bounded away from zero on $(u_1, u_2) \in \mathcal{M}$. Equation (55) gives

$$\frac{\varphi_I(u_1+u_2)}{h(u_1+u_2)} = \frac{\varphi_I(u_1)}{h(u_1)}\frac{\varphi_I(u_2)}{h(u_2)}. \tag{56}$$

Define $\gamma(u) = \frac{\varphi_I(u)}{h(u)}$, which is analytic on an open interval of 0. Equation (56) can be rewritten as

$$\gamma(u_1+u_2) = \gamma(u_1)\gamma(u_2). \tag{57}$$

In Theorem 1 on page 38 of Aczel (1966), it has been shown that the only nonzero analytic solutions of (57) are the exponential functions, $\exp(au)$, where $a \in \mathcal{C}$ is a constant. Namely, $\varphi_I(\frac{s_0}{2}u) = \exp(\tilde{a}u)\psi\left(\frac{\tilde{s}}{2}u\right)$, for some fixed $\tilde{a} \in \mathcal{C}$. Since, for all $u \in \mathbb{R}$, $\varphi_I(-\frac{s_0}{2}u) = \overline{\varphi_I(\frac{s_0}{2}u)}$ and $\psi\left(-\frac{\tilde{s}}{2}u\right) = \overline{\psi\left(\frac{\tilde{s}}{2}u\right)}$, it is straightforward to show $\tilde{a} = if$, for some $f \in \mathbb{R}$. Equivalently,

$$\frac{s_0}{2}I_t = \frac{\tilde{s}}{2}\widetilde{I}_t + f, \tag{58}$$

where the c.f. of I_t is $\varphi_I(u)$, and the c.f. of $\widetilde{I}_t$ is $\psi(u)$. Equation (58) implies the number of points in the support of I_t is also identified. Let the ordered sets $\{m_1, m_2, \cdots, m_l\} \subset \{-k_1, \cdots, 0, \cdots, +k_2\}$ and $\{\widetilde{m}_1, \widetilde{m}_2, \cdots, \widetilde{m}_l\} \subset \{-k_1, \cdots, 0, \cdots, +k_2\}$ denote the supports of I_t and $\widetilde{I}_t$, respectively. Equation (58) implies, for all $i = 1, \cdots, l$,

$$\widetilde{m}_i = \frac{s_0}{\tilde{s}}m_i - f\frac{2}{\tilde{s}}.$$

Since $m_1 = \widetilde{m}_1 = -k_1$, and $m_l = \widetilde{m}_l = +k_2$, $s_0 = \tilde{s}$ and $f = 0$. Therefore, s_0 and the distribution of I_t can be uniquely identified. Then $\varphi_\varepsilon(u) = \frac{\varphi_{\Delta p,1}(u)}{\varphi_I(\frac{s_0}{2}u)\varphi_I(-\frac{s_0}{2}u)}$. □**Proof of Lemma 1**

Equation (43) gives on $\overline{\mathcal{V}}^2$,

$$H(u_1, u_2) := \frac{\varphi_{\Delta I}\left(\frac{s_0}{2}u_1, \frac{s_0}{2}u_2\right)}{\varphi_{\Delta I}\left(\frac{s_0}{2}u_1, 0\right)\varphi_{\Delta I}\left(0, \frac{s_0}{2}u_2\right)}.$$

If the pair $(\widetilde{s}, \psi_{\Delta I}(\cdot,\cdot))$ also satisfies Equation (43), i.e.,

$$H(u_1, u_2) = \frac{\varphi_{\Delta I}\left(\frac{s_0}{2}u_1, \frac{s_0}{2}u_2\right)}{\varphi_{\Delta I}\left(\frac{s_0}{2}u_1, 0\right)\varphi_{\Delta I}\left(0, \frac{s_0}{2}u_2\right)} = \frac{\psi_{\Delta I}\left(\frac{\widetilde{s}}{2}u_1, \frac{\widetilde{s}}{2}u_2\right)}{\psi_{\Delta I}\left(\frac{\widetilde{s}}{2}u_1, 0\right)\psi_{\Delta I}\left(0, \frac{\widetilde{s}}{2}u_2\right)}. \tag{59}$$

Then on $\overline{\mathcal{V}}^2$, which contains a small neighbourhood of $(0, 0)$

$$\varphi_{\Delta I}\left(\frac{s_0}{2}u_1, \frac{s_0}{2}u_2\right)\psi_{\Delta I}\left(\frac{\widetilde{s}}{2}u_1, 0\right)\psi_{\Delta I}\left(0, \frac{\widetilde{s}}{2}u_2\right) = \psi_{\Delta I}\left(\frac{\widetilde{s}}{2}u_1, \frac{\widetilde{s}}{2}u_2\right)\varphi_{\Delta I}\left(\frac{s_0}{2}u_1, 0\right)\varphi_{\Delta I}\left(0, \frac{s_0}{2}u_2\right). \tag{60}$$

Since $\{I_t\}$ is discrete with support $\{-k, \ldots, 0, \ldots, +k\}$, $\varphi_{\Delta I}(\cdot,\cdot)$ and $\psi_{\Delta I}(\cdot,\cdot)$ are analytic. Therefore $\varphi_{\Delta I}(\cdot,\cdot)$ and $\psi_{\Delta I}(\cdot,\cdot)$ have analytic continuations for all complex numbers $(z_1, z_2) \in \mathbb{C}^2$. Furthermore, the analytic continuations, $\varphi_{\Delta I}(z_1, z_2)$ and $\psi_{\Delta I}(z_1, z_2)$ are entire functions, and Equation (60) is satisfied for all $(z_1, z_2) \in \mathbb{C}^2$. Let $\mathbb{Z} \subset \mathbb{C}$ be the set of zeros of $\varphi_{\Delta I}\left(\frac{s_0}{2}\cdot, 0\right)$, and $\widetilde{\mathbb{Z}} \subset \mathbb{C}$ be the set of zeros of $\psi_{\Delta I}\left(\frac{\widetilde{s}}{2}\cdot, 0\right)$. Fix $z_1 = d + fi \in \mathbb{Z}$, where $d, f \in \mathbb{R}$. Then, for any $z \in \mathbb{C}$,

$$\varphi_{\Delta I}\left(\frac{s_0}{2}z_1, \frac{s_0}{2}z\right)\psi_{\Delta I}\left(\frac{\widetilde{s}}{2}z_1, 0\right)\psi_{\Delta I}\left(0, \frac{\widetilde{s}}{2}z\right) = 0 \tag{61}$$

Define $a(z) = \left(\exp\left[iz\frac{s_0}{2}(-2k)\right], \cdots, \exp\left[iz\frac{s_0}{2}(-1)\right], 1, \exp\left[iz\frac{s_0}{2}(1)\right], \cdots, \exp\left[iz\frac{s_0}{2}(2k)\right]\right)^\intercal$, then

$$\varphi_{\Delta I}\left(\frac{s_0}{2}z_1, \frac{s_0}{2}z\right) = a(z)^\intercal \mathcal{Q}_{\Delta I} a(z_1) = a(z)^\intercal A_{\mathcal{Q},\pi} B_{\mathcal{Q}} a(z_1).$$

$z \to \varphi_{\Delta I}\left(\frac{s_0}{2}z_1, \frac{s_0}{2}\cdot\right)$ is the null function if and only if $A_{\mathcal{Q},\pi} B_{\mathcal{Q}} a(z_1) = 0$. Since $A_{\mathcal{Q},\pi}$ is of full column rank, $A_{\mathcal{Q},\pi} B_{\mathcal{Q}} a(z_1) = 0$ if and only if $B_{\mathcal{Q}} a(z_1) = 0$.

$$B_{\mathcal{Q}} a(z_1) = \begin{bmatrix} q_{-k,-k} + q_{-k,-k+1}\exp\left(iz_1\frac{s_0}{2}\right) + q_{-k,-k+2}\exp\left(2iz_1\frac{s_0}{2}\right) + \cdots + q_{-k,k}\exp\left(2kiz_1\frac{s_0}{2}\right) \\ q_{-k+1,-k}\exp\left(-iz_1\frac{s_0}{2}\right) + q_{-k+1,-k+1} + q_{-k+1,-k+2}\exp\left(iz_1\frac{s_0}{2}\right) + \cdots + q_{-k+1,k}\exp\left((2k-1)iz_1\frac{s_0}{2}\right) \\ q_{-k+2,-k}\exp\left(-2iz_1\frac{s_0}{2}\right) + q_{-k+2,-k+1}\exp\left(-iz_1\frac{s_0}{2}\right) + q_{-k+2,-k+2} + \cdots + q_{-k+2,k}\exp\left((2k-2)iz_1\frac{s_0}{2}\right) \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ q_{k-2,-k}\exp\left((-2k+2)iz_1\frac{s_0}{2}\right) + \cdots + q_{k-2,k-2} + q_{k-2,k-1}\exp\left(iz_1\frac{s_0}{2}\right) + q_{k-2,k}\exp\left(2iz_1\frac{s_0}{2}\right) \\ q_{k-1,-k}\exp\left((-2k+1)iz_1\frac{s_0}{2}\right) + q_{k-1,-k+1}\exp\left((-2k+2)iz_1\frac{s_0}{2}\right) + \cdots + q_{k-1,k-1} + q_{k-1,k}\exp\left(iz_1\frac{s_0}{2}\right) \\ q_{k,-k}\exp\left(-2kiz_1\frac{s_0}{2}\right) + q_{k,-k+1}\exp\left((-2k+1)iz_1\frac{s_0}{2}\right) + q_{k,-k+2}\exp\left((-2k+2)iz_1\frac{s_0}{2}\right) + \cdots + q_{k,k} \end{bmatrix}$$

The real part of the first component of $B_{\mathcal{Q}} a(z_1)$ equals to

$$q_{-k,-k} + q_{-k,-k+1}\exp\left(-f\frac{s_0}{2}\right)\cos\frac{ds_0}{2} + q_{-k,-k+2}\exp\left(-2f\frac{s_0}{2}\right)\cos\frac{2ds_0}{2} + \cdots + q_{-k,k}\exp\left(-2kf\frac{s_0}{2}\right)\cos\frac{2kds_0}{2}, \tag{62}$$

while the real part of the last component of $B_{\mathcal{Q}} a(z_1)$ equals to

$$q_{k,-k}\exp\left(2kf\frac{s_0}{2}\right)\cos\frac{2kds_0}{2} + q_{k,-k+1}\exp\left((2k-1)f\frac{s_0}{2}\right)\cos\frac{(2k-1)ds_0}{2} + \cdots + q_{k,k-1}\exp\left(f\frac{s_0}{2}\right)\cos\frac{ds_0}{2} + q_{k,k}. \tag{63}$$

Since $q_{k,k} > 1/2$ and $q_{-k,-k} > 1/2$, either Equation (62) or (63) is strictly larger than zero, no matter what value z_1 takes. Therefore, $A_{\mathcal{Q},\pi} B_{\mathcal{Q}} a(z_1) \neq 0$ and $z \to \varphi_{\Delta I}\left(\frac{s_0}{2}z_1, \frac{s_0}{2}\cdot\right)$ is not the null function. Thus, it is possible to choose $z_2 \in \mathbb{C}$ such that $\varphi_{\Delta I}\left(\frac{s_0}{2}z_1, \frac{s_0}{2}z_2\right) \neq 0$, and $\psi_{\Delta I}\left(0, \frac{\widetilde{s}}{2}z_2\right) \neq 0$. Then Equation (61) leads to $\psi_{\Delta I}\left(\frac{\widetilde{s}}{2}z_1, 0\right) = 0$, therefore $\mathbb{Z} \subset \widetilde{\mathbb{Z}}$. A similar argument shows $\widetilde{\mathbb{Z}} \subset \mathbb{Z}$, therefore $\widetilde{\mathbb{Z}} = \mathbb{Z}$. Since $\varphi_{\Delta I}\left(\frac{s_0}{2}z, 0\right)$ and $\psi_{\Delta I}\left(\frac{\widetilde{s}}{2}z, 0\right)$ have growth order 1, using Hadamard's factorization theorem (see Stein and Shakarchi (2003), Theorem 5.1), we can get that there exists a polynomial R of degree ≤ 1 such that for all $z \in \mathbb{C}$,

$$\varphi_{\Delta I}\left(\frac{s_0}{2}z, 0\right) = \exp(R(z))\psi_{\Delta I}\left(\frac{\widetilde{s}}{2}z, 0\right).$$

Since $\varphi_{\Delta I}(0,0) = \psi_{\Delta I}(0,0) = 1$, there exists a complex number c such that $\varphi_{\Delta I}\left(\frac{s_0}{2}z, 0\right) = \exp(cz)\psi_{\Delta I}\left(\frac{\tilde{s}}{2}z, 0\right)$. Furthermore, for all $z \in \mathbb{R}$, $\varphi_{\Delta I}\left(-\frac{s_0}{2}z, 0\right) = \overline{\varphi_{\Delta I}\left(\frac{s_0}{2}z, 0\right)}$ and $\psi_{\Delta I}\left(-\frac{\tilde{s}}{2}z, 0\right) = \overline{\psi_{\Delta I}\left(\frac{\tilde{s}}{2}z, 0\right)}$. It is straightforward to show $c = if$, for some $f \in \mathbb{R}$. According to the support information, the only possible value of f is zero. Therefore, $\varphi_{\Delta I}\left(\frac{s_0}{2}z, 0\right) = \psi_{\Delta I}\left(\frac{\tilde{s}}{2}z, 0\right)$, for all $z \in \mathbb{C}$. Since $\varphi_{\Delta I}\left(\frac{s_0}{2}z, 0\right) = \varphi_{\Delta I}\left(0, \frac{s_0}{2}z\right)$, and $\psi_{\Delta I}\left(\frac{\tilde{s}}{2}z, 0\right) = \psi_{\Delta I}\left(0, \frac{\tilde{s}}{2}z\right)$, Equation (60) leads to $\varphi_{\Delta I}\left(\frac{s_0}{2}z_1, \frac{s_0}{2}z_2\right) = \psi_{\Delta I}\left(\frac{\tilde{s}}{2}z_1, \frac{\tilde{s}}{2}z_2\right)$, for all $(z_1, z_2) \in \mathbb{C}^2$. Namely, the joint distribution of $\left[\frac{s_0}{2}(I_{t-1} - I_{t-2}), \frac{s_0}{2}(I_t - I_{t-1})\right]$ can be identified by Equation (43). According to the joint support information of $(I_{t-1} - I_{t-2}, I_t - I_{t-1})$, s_0 can be identified. Therefore, $(s_0, \varphi_{\Delta I}(\cdot,\cdot))$ is identified. Then $\varphi_\varepsilon(u) = \frac{\varphi_{\Delta p,1}(u)}{\varphi_{\Delta I}\left(\frac{s_0}{2}u, 0\right)}$. □**Proof of Theorem 8**

According to Lemma 1, s_0 and the joint distribution of $(I_{t-1} - I_{t-2}, I_t - I_{t-1})$ can be identified by Equation (43). For any fixed integer k, $\{I_t\}$ takes values in $\{-k, \cdots, 0, \cdots, +k)$. The probabilities of the first row and the last row of Equation (44) satisfy

$$\Pr(-2k, j) = \pi_k q_{k,-k} q_{-k,-k+j}, \quad \Pr(2k, -j) = \pi_{-k} q_{-k,k} q_{k,k-j}, \quad for\ j = 0, 1, \cdots, 2k$$

where $\Pr(-2k, j)$ and $\Pr(2k, -j)$ are short for $\Pr(I_{t-1} - I_{t-2} = -2k, I_t - I_{t-1} = j)$ and $\Pr(I_{t-1} - I_{t-2} = 2k, I_t - I_{t-1} = -j)$, respectively. Along with the fact that $\sum_{j=0}^{2k} q_{-k,-k+j} = 1$ and $\sum_{j=0}^{2k} q_{k,k-j} = 1$, π_k, π_{-k}, $Q_{k,\cdot}$ and $Q_{-k,\cdot}$ can be identified. The probabilities of the second row and the second last row of Equation (44) satisfy

$$\Pr(-2k+1, -1) = \pi_k q_{k,-k+1} q_{-k+1,-k}, \quad \Pr(-2k+1, 2k) = \pi_{k-1} q_{k-1,-k} q_{-k,k}, \tag{64}$$

$$\Pr(-2k+1, j) = \pi_k q_{k,-k+1} q_{-k+1,-k+j+1} + \pi_{k-1} q_{k-1,-k} q_{-k,-k+j}, \quad for\ j = 0, 1, \cdots, 2k-1 \tag{65}$$

$$\Pr(2k-1, 1) = \pi_{-k} q_{-k,k-1} q_{k-1,k}, \quad \Pr(2k-1, -2k) = \pi_{-k+1} q_{-k+1,k} q_{k,-k}, \tag{66}$$

$$\Pr(2k-1, -j) = \pi_{-k} q_{-k,k-1} q_{k-1,k-1-j} + \pi_{-k+1} q_{-k+1,k} q_{k,k-j}, \quad for\ j = 0, 1, \cdots, 2k-1 \tag{67}$$

Equation (64) and (66) can be used to identify $\pi_{k-1} \times q_{k-1,-k}$, $\pi_{-k+1} \times q_{-k+1,k}$, $q_{k-1,k}$ and $q_{-k+1,-k}$. Then Equation (65) and (67) can be used to identify $q_{-k+1,j}$, for $j = -k+1, \cdots, k$ and $q_{k-1,j}$, for $j = -k, \cdots, k-1$, respectively. Consequently, π_{k-1} and π_{-k+1} can be identified. Following the same strategy, the probabilities of the third row and the third last row of Equation (44) can be used to identify π_{k-2}, π_{-k+2}, $Q_{k-2,\cdot}$ and $Q_{-k+2,\cdot}$. Essentially, the same strategy can be applied recursively to identify π and $\mathcal{Q}$. □

B. Proofs for Estimation Results in Section 3

We start with some background material. The complex conjugate and the real and imaginary parts of a complex number z are denoted by $\overline{z}$, $\mathrm{Re}(z)$, and $\mathrm{Im}(z)$, respectively. Denote the joint characteristic function and empirical c.f. of $\left(\Delta p_t, \cdots, \Delta p_{t-(d-1)}\right)$, for $d \geq 1$, by

$$\varphi_{\Delta p,d}\left(u_1, u_2, \cdots, u_d\right) = \mathbb{E}\left(\exp\left(iu_1\Delta p_t + iu_2\Delta p_{t-1} + \cdots + iu_d\Delta p_{t-(d-1)}\right)\right), \tag{68}$$

$$\varphi_{T,d}\left(u_1, u_2, \cdots, u_d\right) = \frac{1}{T}\sum_{t=1}^{T}\exp\left(iu_1\Delta p_t + iu_2\Delta p_{t-1} + \cdots + iu_d\Delta p_{t-(d-1)}\right), \tag{69}$$

where $\mathbb{E}\left(\varphi_{T,d}\left(u_1, u_2, \cdots, u_d\right)\right) = \varphi_{\Delta p,d}\left(u_1, u_2, \cdots, u_d\right)$. Define the empirical characteristic function process

$$W_{T,2}\left(u, u'\right) := \sqrt{T}\left(\varphi_{T,2}\left(u, u'\right) - \varphi_{\Delta p,2}\left(u, u'\right)\right) \tag{70}$$

Concerning the covariance structure, since $\{\Delta p_t\}_{t=1}^{T}$ is strictly stationary and 1-dependent, we are able to show that

$$\begin{aligned}\lim_{T\to\infty} Cov\left(W_{T,2}\left(u, u'\right), W_{T,2}\left(v, v'\right)\right) &= \sum_{s=-2}^{2} Cov\left[\exp\left(iu\Delta p_t + iu'\Delta p_{t-1}\right), \exp\left(iv\Delta p_{t-s} + iv'\Delta p_{t-s-1}\right)\right] \\ &= \varphi_{\Delta p,4}\left(-v, -v', u, u'\right) + \varphi_{\Delta p,4}\left(u, u', -v, -v'\right) + \varphi_{\Delta p,3}\left(-v, u - v', u'\right) + \varphi_{\Delta p,3}\left(u, u' - v, -v'\right) \\ &\quad + \varphi_{\Delta p,2}\left(u - v, u' - v'\right) - 5\varphi_{\Delta p,2}\left(u, u'\right)\varphi_{\Delta p,2}\left(-v, -v'\right),\end{aligned} \tag{71}$$

which can be consistently estimated by the sample analogs. Given $\{\Delta p_t\}_{t=1}^{T}$ is strictly stationary and 1-dependent, following Theorem 2.1 and Theorem 2.2 of Feuerverger (1990), we have the following result:

Lemma 2. *For $d \geq 1$,*

(i) $\varphi_{T,d}\left(u_1, u_2, \cdots, u_d\right) \to \varphi_{\Delta p,d}\left(u_1, u_2, \cdots, u_d\right)$ *a.s.,* $\forall\left(u_1, u_2, \cdots, u_d\right) \in \mathbb{R}^d$*;*

(ii) $\sup_{|u_j|\leq M,\ j=1,\cdots,d}\left|\varphi_{T,d}\left(u_1, u_2, \cdots, u_d\right) - \varphi_{\Delta p,d}\left(u_1, u_2, \cdots, u_d\right)\right| \to 0$ *a.s., for any fixed* $0 < M < \infty$*;*

(iii) $\forall k, \quad \forall\left(u_1, u_1'\right), \cdots, \left(u_k, u_k'\right) \in \mathbb{R}^2$:

$$\left(W_{T,2}\left(u_1, u_1'\right), \cdots, W_{T,2}\left(u_k, u_k'\right)\right) \to^d \left(W_2\left(u_1, u_1'\right), \cdots, W_2\left(u_k, u_k'\right)\right),$$

where $\{W_2\left(u, u'\right) = \overline{W_2\left(-u, -u'\right)}, \left(u, u'\right) \in \mathbb{R}^2\}$ *is a zero mean complex-valued Gaussian process with the covariance structure (71).*

Proof of Theorem 3 Under Assumption 2′, $\mathcal{S}$ is compact and $\mathcal{U}$ contains only finite number of grid points. It is straightforward to show $\sup_{s\in\mathcal{S}} |J_T(s,\mathcal{U}) - J(s,\mathcal{U})| = o_p(1)$, due to the fact that the characteristic functions are bounded by unity in norm and Lemma 2(ii). In addition, the identification result in Theorem 2 establishes $J(s,\mathcal{U}) > 0$, for $s \in \mathcal{S}, s \neq s_0$. Then using Theorem 2.1 of Newey and McFadden (1994), the consistency of $\widehat{s}_{ecf}$ is established. Assumptions 1 and 2′ assure that $\inf_{(u,u')\in\mathcal{U}} |\varphi_{\Delta p,1}(u)\varphi_{\Delta p,1}(u')| > 0$, or equivalently $\sup_{(u,u')\in\mathcal{U}} |\frac{1}{\varphi_{\Delta p,1}(u)\varphi_{\Delta p,1}(u')}| < \infty$. Along with Lemma 2 and the property of characteristic functions, we are able to show $\sup_{(u,u')\in\mathcal{U}} |H_T(u,u') - H(u,u')| = o_p(1)$. Also, $H(u,u') = R(u,u';s_0)$ is uniformly bounded on $\mathcal{U}$. Then the consistency of $\widehat{s}_{ecf,2}$ can be established similarly. Assumptions 1 and 2′ also ensure that

$$\min_{(u,u')\in\mathcal{U}} \left(|\varphi^2_{\Delta p,1}(u)|, |\varphi^2_{\Delta p,1}(u')|, |\varphi_{\Delta p,2}(u,u)|, |\varphi_{\Delta p,2}(u',u')|\right) \geq \delta_1 > 0,$$

for some constant δ_1. This establishes that, as stated in Remark 1, if the cutoff c in the construction of $\mathcal{U}$ is chosen small enough ($c < \delta_1$), then the corresponding trimming constraints $\mathcal{I}\left\{|\varphi^2_{T,1}(u)| > c\right\}$ and $\mathcal{I}\{|\varphi_{T,2}(u,u)| > c\}$ are not binding asymptotically. □Before we prove Theorem 4 on the asymptotic normality, we define quantities that appear inside the limiting variances of our estimators. Let $W_2(u,u')$ be a Gaussian process with covariance function

$$\begin{aligned} &\operatorname{cov}\left(W_2(u,u'), W_2(v,v')\right) \\ &= \varphi_{\Delta p,4}(-v,-v',u,u') + \varphi_{\Delta p,4}(u,u',-v,-v') + \varphi_{\Delta p,3}(-v,u-v',u') \\ &+ \varphi_{\Delta p,3}(u,u'-v,-v') + \varphi_{\Delta p,2}(u-v,u'-v') - 5\varphi_{\Delta p,2}(u,u')\varphi_{\Delta p,2}(-v,-v'), \end{aligned} \tag{72}$$

where the joint c.f. of $(\Delta p_t, \ldots, \Delta p_{t-(d-1)})$, for $d \geq 1$, is denoted by $\varphi_{\Delta p,d}(u_1,u_2,\ldots,u_d) = \mathbb{E}\left(\exp\left(iu_1\Delta p_t + iu_2\Delta p_{t-1} + \cdots + iu_d\Delta p_{t-(d-1)}\right)\right)$. Define the real random variables

$$G(u,u') := \operatorname{Re}\left(\frac{W_2(u,u')}{\varphi_{\Delta p,1}(u)\varphi_{\Delta p,1}(u')}\right) - R(u,u';s_0)\operatorname{Re}\left(\frac{W_2(0,u')}{\varphi_{\Delta p,1}(u')}\right) - R(u,u';s_0)\operatorname{Re}\left(\frac{W_2(u,0)}{\varphi_{\Delta p,1}(u)}\right) \tag{73}$$

$$= \frac{1}{2}\left[\frac{W_2(u,u')}{\varphi_{\Delta p,1}(u)\varphi_{\Delta p,1}(u')} + \frac{W_2(-u,-u')}{\varphi_{\Delta p,1}(-u)\varphi_{\Delta p,1}(-u')}\right] - \frac{R(u,u';s_0)}{2}\left[\frac{W_2(0,u')}{\varphi_{\Delta p,1}(u')} + \frac{W_2(0,-u')}{\varphi_{\Delta p,1}(-u')} + \frac{W_2(u,0)}{\varphi_{\Delta p,1}(u)} + \frac{W_2(-u,0)}{\varphi_{\Delta p,1}(-u)}\right].$$

Denote the vectorized version of $\{G(u,u'); \forall (u,u') \in \mathcal{U}\}$ as $G(\mathcal{U})$. Let the covariance of $G(\mathcal{U})$ be Σ_0, which is a $|\mathcal{U}| \times |\mathcal{U}|$ matrix. Every component of Σ_0, $\operatorname{cov}(G(u,u'), G(v,v'))$, can be calculated using (72). We provide a detailed instruction on how to calculate Σ_0 and a consistent estimator $\widehat{\Sigma}_0$

in Appendix C. **Proof of Theorem 4** In the proof, we shall verify the conditions in Theorem 3.1 of Newey and McFadden (1994) for the asymptotic normality. In the following ∇_s and ∇_{ss} denote the first and second derivatives of a function with respect to s, respectively. We first derive a few useful formulae.

$$\nabla_s R(u,u';s) = \frac{\frac{u'}{2}\sin\left(u\frac{s}{2}\right)\cos\left(u\frac{s}{2}\right)+\frac{u}{2}\sin\left(u'\frac{s}{2}\right)\cos\left(u'\frac{s}{2}\right)}{\left[\cos\left(u\frac{s}{2}\right)\cos\left(u'\frac{s}{2}\right)\right]^2}, \tag{74}$$

$$\nabla_{ss} R(u,u';s) = \frac{\frac{uu'}{2}\cos\left(u\frac{s}{2}\right)\cos\left(u'\frac{s}{2}\right)+\sin\left(u\frac{s}{2}\right)\sin\left(u'\frac{s}{2}\right)\left(\frac{u'^2}{2}\cos^2\left(u\frac{s}{2}\right)+\frac{u^2}{2}\cos^2\left(u'\frac{s}{2}\right)\right)}{\left[\cos\left(u\frac{s}{2}\right)\cos\left(u'\frac{s}{2}\right)\right]^3}.$$

Note that on the boundary point 0, $\nabla_s R(u,u';0)\equiv 0$, $\forall (u,u')\in\mathcal{U}$. Under Assumption 2′ (ii),

$$\sup_{(u,u')\in\mathcal{U},s\in\mathcal{S}} |\nabla_s R(u,u';s)| < M_3, \qquad \sup_{(u,u')\in\mathcal{U},s\in\mathcal{S}} |\nabla_{ss} R(u,u';s)| < M_4, \tag{75}$$

for some fixed number $0 < M_3, M_4 < \infty$.

(i) We have

$$\nabla_{ss} J\left(s_0,\mathcal{U}\right) = 2\sum_{(u,u')\in\mathcal{U}} |\varphi_{\Delta p,1}(u)|^2|\varphi_{\Delta p,1}(u')|^2\ \nabla_s^2 R(u,u';s_0) > 0 \tag{76}$$

$$\nabla_s J_T\left(s,\mathcal{U}\right) = -2\sum_{(u,u')\in\mathcal{U}} \nabla_s R(u,u';s)\,\mathrm{Re}\left(\overline{\varphi_{T,1}(u)\varphi_{T,1}(u')}\left[\varphi_{T,2}(u,u')-\varphi_{T,1}(u)\varphi_{T,1}(u')R(u,u';s)\right]\right) \tag{77}$$

$$\begin{aligned}\nabla_{ss} J_T\left(s,\mathcal{U}\right) = &-2\sum_{(u,u')\in\mathcal{U}} \nabla_{ss} R(u,u';s)\,\mathrm{Re}\left(\overline{\varphi_{T,1}(u)\varphi_{T,1}(u')}\left[\varphi_{T,2}(u,u')-\varphi_{T,1}(u)\varphi_{T,1}(u')R(u,u';s)\right]\right)\\ &+2\sum_{(u,u')\in\mathcal{U}} |\varphi_{T,1}(u)|^2|\varphi_{T,1}(u')|^2\ \nabla_s^2 R(u,u';s)\end{aligned}$$

Due to Assumption 2′ (ii), Equation (75), Lemma 2(ii), and the properties of characteristic functions, it is straightforward to show $\sup_{s\in\mathcal{S}}\left|\nabla_{ss} J_T\left(s,\mathcal{U}\right)-\nabla_{ss} J\left(s,\mathcal{U}\right)\right| = o_p(1)$. Furthermore, $\nabla_{ss} J\left(s_0,\mathcal{U}\right) > 0$. $\widehat{s}_{ecf}$ satisfies the first order condition $\nabla_s J_T\left(\widehat{s}_{ecf},\mathcal{U}\right) = 0$. Expanding the first order condition around s_0, we have

$$0 = \nabla_s J_T\left(\widehat{s}_{ecf},\mathcal{U}\right) = \nabla_s J_T\left(s_0,\mathcal{U}\right) + \nabla_{ss} J_T\left(s_0,\mathcal{U}\right)\left[\widehat{s}_{ecf}-s_0\right] + o_p(1/\sqrt{T})$$

$$\begin{aligned}\sqrt{T}\left[\widehat{s}_{ecf}-s_0\right] &= \frac{-\sqrt{T}\nabla_s J_T\left(s_0,\mathcal{U}\right)}{\nabla_{ss} J\left(s_0,\mathcal{U}\right)} + o_p(1)\\ &= \frac{\sum_{(u,u')\in\mathcal{U}} \nabla_s R(u,u';s_0)\,\mathrm{Re}\left(\overline{\varphi_{\Delta p,1}(u)\varphi_{\Delta p,1}(u')}\sqrt{T}\left[\varphi_{T,2}(u,u')-\varphi_{T,1}(u)\varphi_{T,1}(u')R(u,u';s_0)\right]\right)}{\sum_{(u,u')\in\mathcal{U}} |\varphi_{\Delta p,1}(u)|^2|\varphi_{\Delta p,1}(u')|^2\ \nabla_s^2 R(u,u';s_0)} + o_p(1)\end{aligned}$$

where

$$\sqrt{T}\left[\varphi_{T,2}(u,u')-\varphi_{T,1}(u)\varphi_{T,1}(u')R(u,u';s_0)\right]=$$
$$\sqrt{T}\begin{bmatrix}\varphi_{T,2}(u,u')-\varphi_{\Delta p,2}(u,u')\\ -\varphi_{\Delta p,1}(u)R(u,u';s_0)\left[\varphi_{T,1}(u')-\varphi_{\Delta p,1}(u')\right]\\ -\varphi_{T,1}(u')R(u,u';s_0)\left[\varphi_{T,1}(u)-\varphi_{\Delta p,1}(u)\right]\end{bmatrix}$$
$$\to^d W_2\left(u,u'\right)-R(u,u';s_0)\varphi_{\Delta p,1}(u)W_2\left(0,u'\right)-R(u,u';s_0)\varphi_{\Delta p,1}(u')W_2\left(u,0\right)$$

The last convergence result follows from Lemma 2 (iii). Therefore,

$$\sqrt{T}\left[\widehat{s}_{ecf}-s_0\right]\to^d\frac{\sum_{(u,u')\in\mathcal{U}}\nabla_s R(u,u';s_0)|\varphi_{\Delta p,1}(u)|^2|\varphi_{\Delta p,1}(u')|^2G\left(u,u'\right)}{\sum_{(u,u')\in\mathcal{U}}|\varphi_{\Delta p,1}(u)|^2|\varphi_{\Delta p,1}(u')|^2\ \nabla_s^2R(u,u';s_0)}$$

(ii) We derive the asymptotic normality of $\widehat{s}_{ecf,2}$ in a similar manner.

$$\nabla_{ss}Q\left(s_0,\mathcal{U}\right)=2\sum_{(u,u')\in\mathcal{U}}\nabla_s^2R(u,u';s_0)>0$$
$$\nabla_sQ_T\left(s,\mathcal{U}\right)=-2\sum_{(u,u')\in\mathcal{U}}\nabla_sR(u,u';s)\left(\mathrm{Re}\left[H_T(u,u')\right]-R(u,u';s)\right)$$
$$\nabla_{ss}Q_T\left(s,\mathcal{U}\right)=-2\sum_{(u,u')\in\mathcal{U}}\nabla_{ss}R(u,u';s)\left(\mathrm{Re}\left[H_T(u,u')\right]-R(u,u';s)\right)+2\sum_{(u,u')\in\mathcal{U}}\nabla_s^2R(u,u';s)$$

Due to Assumption 2′ (ii), Equation (75), Lemma 2(ii), and the properties of characteristic functions, it is easy to show $\sup_{s\in\mathcal{S}}|\nabla_{ss}Q\left(s,\mathcal{U}\right)|<\infty$ and $\sup_{s\in\mathcal{S}}|\nabla_{ss}Q_T\left(s,\mathcal{U}\right)-\nabla_{ss}Q\left(s,\mathcal{U}\right)|=o_p(1)$. Furthermore, $\nabla_{ss}Q\left(s_0,\mathcal{U}\right)>0$. $\widehat{s}_{ecf,2}$ satisfies the first order condition $\nabla_sQ_T\left(\widehat{s}_{ecf,2},\mathcal{U}\right)=0$. Expanding the first order condition around s_0, similar to part (i), we can get

$$\sqrt{T}\left[\widehat{s}_{ecf,2}-s_0\right]\to^d\frac{\sum_{(u,u')\in\mathcal{U}}\nabla_sR(u,u';s_0)G\left(u,u'\right)}{\sum_{(u,u')\in\mathcal{U}}\nabla_s^2R(u,u';s_0)}.$$

□Proof of Theorem 5

The conditions in Theorem 3.2 of Newey and McFadden (1994) can be easily verified. We have

$$\nabla_{ss}Q_D\left(s_0,\mathcal{U}\right)=2\nabla_sR(\mathcal{U};s_0)^\intercal D\nabla_sR(\mathcal{U};s_0)>0$$
$$\nabla_sQ_{\widehat{D}_T,T}\left(s,\mathcal{U}\right)=-2\left[\mathrm{Re}\left(H_T(\mathcal{U})\right)-R(\mathcal{U};s)\right]^\intercal\widehat{D}_T\nabla_sR(\mathcal{U};s),$$
$$\nabla_{ss}Q_{\widehat{D}_T,T}\left(s_0,\mathcal{U}\right)=-2\left[\mathrm{Re}\left(H_T(\mathcal{U})\right)-R(\mathcal{U};s_0)\right]^\intercal\widehat{D}_T\nabla_{ss}R(\mathcal{U};s_0)+2\nabla_sR(\mathcal{U};s_0)^\intercal\widehat{D}_T\nabla_sR(\mathcal{U};s_0)$$
$$\to^p\nabla_{ss}Q_D\left(s_0,\mathcal{U}\right).$$

$\widehat{s}_{ecf,\widehat{D}_T}$ satisfies the first order condition $\nabla_s Q_{\widehat{D}_T,T}\left(\widehat{s}_{ecf,\widehat{D}_T},\mathcal{U}\right) = 0$. Expanding the first order condition around s_0, similar to the proof of Theorem 4, we can obtain

$$\sqrt{T}\left[\widehat{s}_{ecf,\widehat{D}_T} - s_0\right] \to^d \frac{\nabla_s R(\mathcal{U};s_0)^\intercal D \times G(\mathcal{U})}{\nabla_s R(\mathcal{U};s_0)^\intercal D \nabla_s R(\mathcal{U};s_0)}.$$

The asymptotic variance of $\widehat{s}_{ecf,\widehat{D}_T}$ is given by

$$(\nabla_s R(\mathcal{U};s_0)^\intercal D \nabla_s R(\mathcal{U};s_0))^{-1} \times \nabla_s R(\mathcal{U};s_0)^\intercal D \Sigma_0 D \nabla_s R(\mathcal{U};s_0) \times (\nabla_s R(\mathcal{U};s_0)^\intercal D \nabla_s R(\mathcal{U};s_0))^{-1} \tag{78}$$

(ii) Follows since (78) is minimized when $D = \Sigma_0^{-1}$. The asymptotic variance of the optimal estimator $\widehat{s}^*_{ecf}$ equals $\left(\nabla_s R(\mathcal{U};s_0)^\intercal \Sigma_0^{-1} \nabla_s R(\mathcal{U};s_0)\right)^{-1}$. □

C. How to Calculate Σ_0

In this section we provide a detailed construction of Σ_0. Recall that Σ_0 denotes the covariance matrix of $G(\mathcal{U})$, where $G(\mathcal{U})$ denotes the vectorized version of $\{G(u,u');\forall(u,u')\in\mathcal{U}\}$, and

$$G(u,u') = \frac{1}{2}\left[\frac{W_2(u,u')}{\varphi_{\Delta p,1}(u)\varphi_{\Delta p,1}(u')} + \frac{W_2(-u,-u')}{\varphi_{\Delta p,1}(-u)\varphi_{\Delta p,1}(-u')}\right] - \frac{R(u,u';s_0)}{2}\left[\frac{W_2(0,u')}{\varphi_{\Delta p,1}(u')} + \frac{W_2(0,-u')}{\varphi_{\Delta p,1}(-u')} + \frac{W_2(u,0)}{\varphi_{\Delta p,1}(u)} + \frac{W_2(-u,0)}{\varphi_{\Delta p,1}(-u)}\right].$$

The construction of Σ_0 depends on the chosen vectorization for $\mathcal{U}$. One possibility, which we use in this section, is given by $[(u_1,u_1'),(u_2,u_1'),\ldots,(u_n,u_1'),(u_1,u_2'),(u_2,u_2'),\ldots,(u_n,u_m')]^\intercal$, where $u_1,\ldots,u_n$ and $u_1',\ldots,u_m'$ denote the elements (in increasing order) along the first and second dimension of $\mathcal{U}$, respectively. For simplicity, we also assume $m=n$ and $u_i = u_i'$, $\forall i = 1,\ldots,n$, i.e., $\mathcal{U} = (u_1,\ldots,u_n)\times(u_1,\ldots,u_n)$. To avoid a singular Σ_0, it is important to choose $u_1 > 0$. However, for the computation of $G(u,u')$ we need the elements of the form $W_2(u_i,0)$ and $W_2(-u_i,0) = \overline{W_2(u_i,0)}$ (recall that $W_2(0,u) = W_2(u,0)$, $\forall u$; this is also the reason why including $u_1 = 0$ as a grid point, together with the construction of $\mathcal{U}$ as a cross product, will lead to a singular Σ_0). Hence we augment the vectorization of $\mathcal{U}$ to $[(u_1,0),\ldots,(u_n,0),(u_1,u_1'),(u_2,u_1'),\ldots,(u_n,u_1'),(u_1,u_2'),(u_2,u_2'),\ldots,(u_n,u_n')]^\intercal$. Denote the corresponding vectorizations of $\{W_2(u,0),W_2(u,u');\forall(u,u')\in\mathcal{U}\}$ and $\{W_2(-u,0),W_2(-u,-u');\forall(u,u')\in\mathcal{U}\}$ by $W_2(\mathcal{U})$ and $W_2(-\mathcal{U})$, respectively. Also denote the stacked $2(|\mathcal{U}|+n)$ dimensional vector $[W_2(\mathcal{U})^\intercal, W_2(-\mathcal{U})^\intercal]^\intercal$ by $\overline{W_2}(\mathcal{U})$. The elements of the $2(|\mathcal{U}|+n)\times 2(|\mathcal{U}|+n)$ covariance matrix of $\overline{W_2}(\mathcal{U})$ are given by (71), which applies to any $(u,u'),(v,v')\in\mathbb{R}^2$:

$$Var(\overline{W_2}(\mathcal{U})) = \begin{bmatrix} Var(W_2(\mathcal{U})), & Cov(W_2(\mathcal{U}),W_2(-\mathcal{U})) \\ Cov(W_2(-\mathcal{U}),W_2(\mathcal{U})), & Var(W_2(-\mathcal{U})) \end{bmatrix}. \tag{79}$$

Next, we define several matrices:

- First, the two $|\mathcal{U}| \times |\mathcal{U}|$ diagonal matrices

$$\Phi_1 := diag\left(\mathbf{1}_{n\times 1} \otimes [\varphi_{\Delta p,1}(u_1), \ldots, \varphi_{\Delta p,1}(u_n)]^\intercal\right), \tag{80}$$

$$\Phi_2 := diag\left([\varphi_{\Delta p,1}(u_1), \ldots, \varphi_{\Delta p,1}(u_n)]^\intercal \otimes \mathbf{1}_{n\times 1}\right), \tag{81}$$

  where $\mathbf{1}_{n\times 1}$ denotes a column vector of ones of length n and $\otimes$ signifies the Kronecker product.

- Second, the two $|\mathcal{U}| \times (|\mathcal{U}| + n)$ matrices

$$D_1 := \left[\mathbf{0}_{|\mathcal{U}|\times n}\,,\ \frac{1}{2}\left(\Phi_1\Phi_2\right)^{-1}\right], \qquad D_2 := \left[\mathbf{0}_{|\mathcal{U}|\times n}\,,\ \frac{1}{2}\left(\overline{\Phi_1\Phi_2}\right)^{-1}\right], \tag{82}$$

  where $\mathbf{0}_{|\mathcal{U}|\times n}$ denotes a matrix of zeros of dimension $|\mathcal{U}| \times n$, and where we use the fact that $\overline{\varphi_{\Delta p,1}(u)} = \varphi_{\Delta p,1}(-u)$, $\forall u$.

- Third, the two $|\mathcal{U}| \times (|\mathcal{U}| + n)$ matrices (note that $|\mathcal{U}| = n^2$)

$$M_{3,4} := \left[I_{n\times n} \otimes \mathbf{1}_{n\times 1}\,,\ \mathbf{0}_{|\mathcal{U}|\times|\mathcal{U}|}\right], \qquad M_{5,6} := \left[\mathbf{1}_{n\times 1} \otimes I_{n\times n}\,,\ \mathbf{0}_{|\mathcal{U}|\times|\mathcal{U}|}\right], \tag{83}$$

  where $I_{n\times n}$ denotes an $n \times n$ identity matrix.

- Fourth, the $|\mathcal{U}| \times |\mathcal{U}|$ matrix

$$R := -\frac{1}{2} diag\left(R(\mathcal{U}; s_0)\right), \tag{84}$$

  where $R(\mathcal{U}; s_0)$ represents the (unaugmented) vectorization of $\{R(u, u'; s_0); \forall (u, u') \in \mathcal{U}\}$.

- And finally, the four $|\mathcal{U}| \times (|\mathcal{U}| + n)$ matrices

$$D_3 := R\,\Phi_2\,M_{3,4}\,, \qquad D_4 := R\,\overline{\Phi_2}\,M_{3,4}\,, \qquad D_5 := R\,\Phi_1\,M_{3,4}\,, \qquad D_6 := R\,\overline{\Phi_1}\,M_{3,4}\,. \tag{85}$$

With the above matrices we can calculate $G(\mathcal{U})$ as

$$\begin{aligned} G(\mathcal{U}) &= D_1W_2(\mathcal{U}) + D_2W_2(-\mathcal{U}) + D_3W_2(\mathcal{U}) + D_4W_2(-\mathcal{U}) + D_5W_2(\mathcal{U}) + D_6W_2(-\mathcal{U}) \\ &= (D_1 + D_3 + D_5)W_2(\mathcal{U}) + (D_2 + D_4 + D_6)W_2(-\mathcal{U}) \\ &= [D_1 + D_3 + D_5\,,\ D_2 + D_4 + D_6]\,\overline{W_2}(\mathcal{U}) \begin{bmatrix} D_1^\intercal + D_3^\intercal + D_5^\intercal \\ D_2^\intercal + D_4^\intercal + D_6^\intercal \end{bmatrix}. \end{aligned}$$

Thus, the covariance matrix Σ_0 is given by

$$\begin{aligned}\Sigma_0 &= Var(G(\mathcal{U})) \\ &= [D_1 + D_3 + D_5\,,\, D_2 + D_4 + D_6]\, Var(\overline{W_2}(\mathcal{U})) \begin{bmatrix} D_1^\intercal + D_3^\intercal + D_5^\intercal \\ D_2^\intercal + D_4^\intercal + D_6^\intercal \end{bmatrix}. \end{aligned} \tag{86}$$

To obtain an estimate $\widehat{\Sigma}_0$ of Σ_0, simply replace the unknown population quantities in (79), (80), (81) and (84) by their sample counterparts. For (79), the individual sample covariances can be computed from (71).

D. Additional Figures

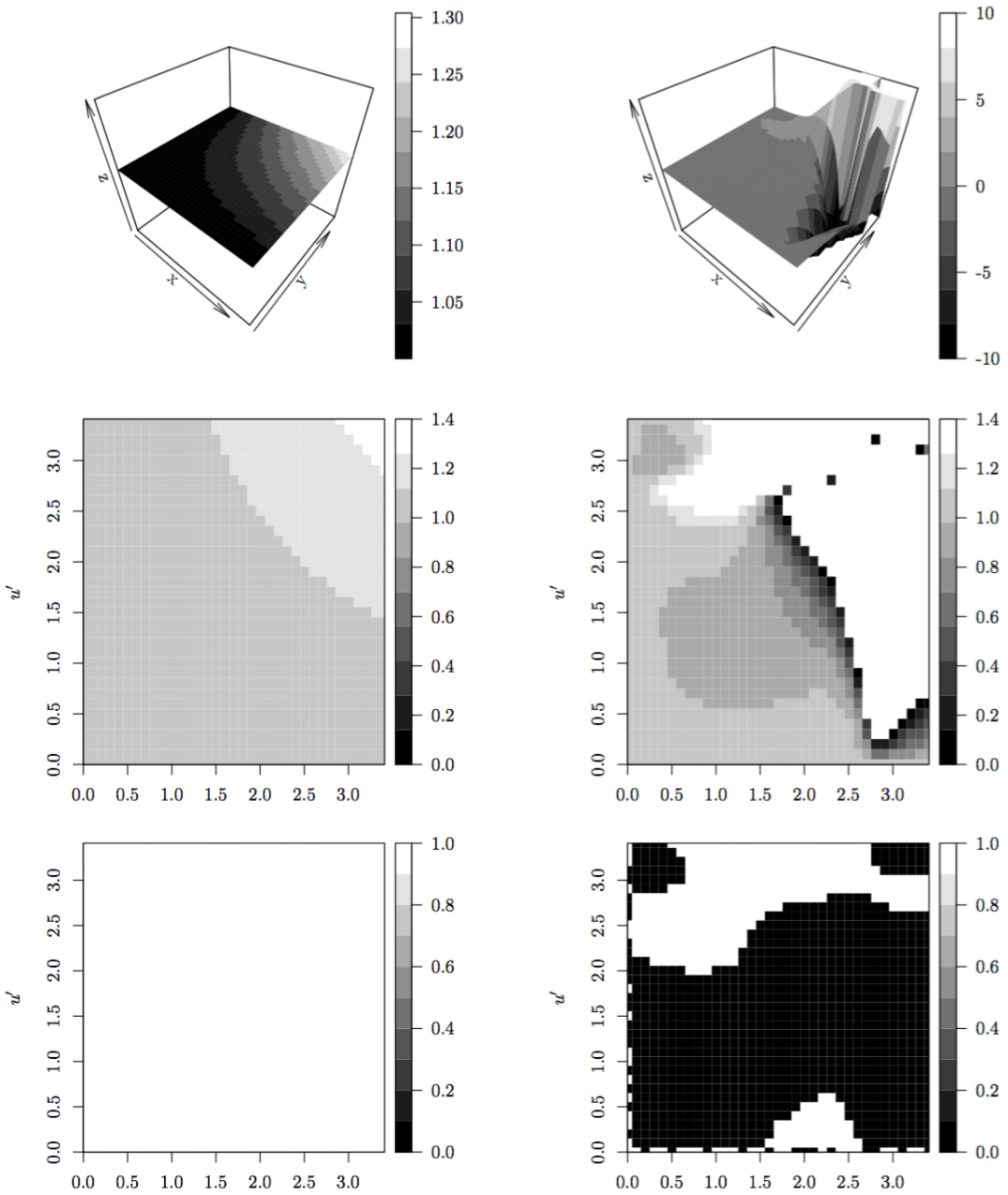

Figure 7: An example of the population quantity $H(u, u')$ and a (simulated) realization of the empirical counterpart $H_T(u, u')$, for $s_0 = 0.3$, $n_g = 20$, $c = 0.1$, $\varepsilon_t \sim \mathrm{N}(0, 1)$, and $T = 250$. The left column pertains to H and shows (from top to bottom) a surface plot, a heat map, and a "binary" heat map that takes the value 1 if $H(u, u') \geq 1$; the right column exhibits the equivalents for H_T.

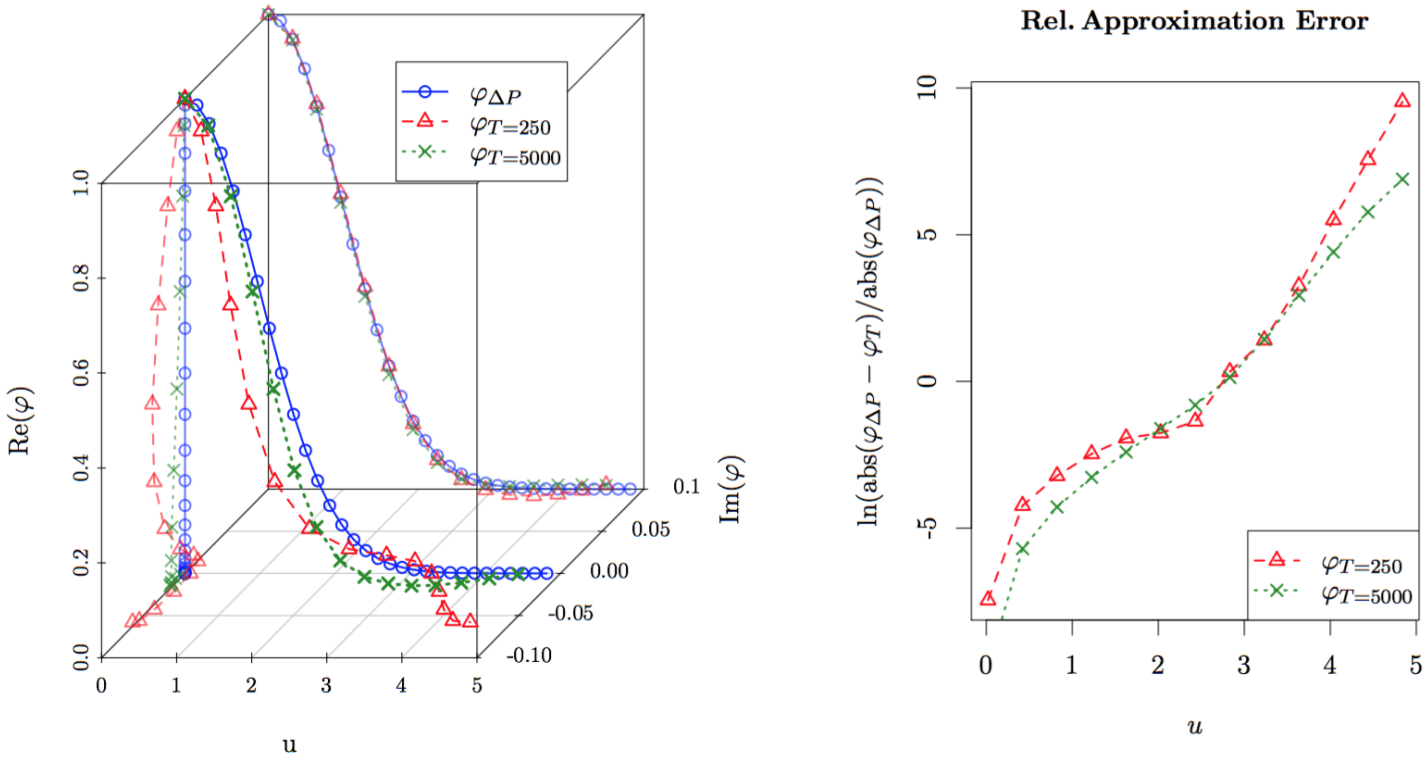

Figure 8: True c.f. $\varphi_{\Delta P}$ and two corresponding e.c.f.'s $\varphi_{T=250}$ and $\varphi_{T=5000}$, along with the corresponding approximation errors. The real and imaginary part of $\varphi_{T=250}$ and $\varphi_{T=5000}$ in the left plot are projected onto the respective planes. The e.c.f.'s are computed from a simulated sample of length $T = 5,000$ with $s_0 = 0.3$ and $\varepsilon_t \sim \mathrm{N}(0,1)$, where $\varphi_{T=250}$ is computed from the first 250 elements of the sample.

References

Aczel, J. (1966). *Lectures on Functional Equations and Their Applications.* Academic Press, New York.

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

Angel, J. J., Harris, L. E., and Spatt, C. S. (2011). Equity trading in the 21st century. *The Quarterly Journal of Finance*, 1(01):1–53.

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.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2:107–144.

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.

Carroll, R., Ruppert, D., Stefanski, L., and Crainiceanu, C. (2006). *Measurement error in nonlinear models.* Chapman and Hall.

Castura, J., Litzenberger, R., Gorelick, R., and Dwivedi, Y. (2010). Market efficiency and microstructure evolution in us equity markets: A high-frequency perspective. *Russell: The Journal Of The Bertrand Russell Archives*, 24.

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.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 187–220.

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*, page hhs053.

Feuerverger, A. (1990). An efficiency result for the empirical characteristic function in stationary time-series models. *Canadian Journal of Statistics*, 18(2):155–161.

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

Gassiat, E. and Rousseau, J. (2016). Nonparametric finite translation hidden markov models and extensions. *Bernoulli*, 22:193 – 212.

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.

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.

Hu, Y. (2016). Microeconomic models with latent variables: Applications of measurement error models in empirical industrial organization and labor economics. *Journal of Econometrics, forthcoming.*

Jones, C. M. (2002). A century of stock market liquidity and trading costs. *Available at SSRN 313681.*

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.

Newey, W. and McFadden, D. (1994). *Handbook of Econometrics*, volume IV, chapter 36 Large Sample Estimation and Hypothesis Testing. Elsevier.

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.

Stein, E. and Shakarchi, R. (2003). *Complex Analysis*. Princeton, NJ: Princeton University Press.

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.

Zhang, H. and Hodges, S. (2012). An extended model of effective bid-ask spread.