Dynamic Autoregressive Liquidity (DArLiQ)

Christian M. Hafner[*a], Oliver B. Linton[†b], and Linqi Wang[‡b]

[a]Louvain Institute of Data Analysis and Modelling in economics and statistics (LIDAM), and ISBA, UCLouvain

[b]Faculty of Economics, University of Cambridge

June 28, 2023

Abstract

We introduce a new class of semiparametric dynamic autoregressive models for the Amihud illiquidity measure, which captures both the long-run trend in the illiquidity series with a nonparametric component and the short-run dynamics with an autoregressive component. We develop a generalized method of moments (GMM) estimator based on conditional moment restrictions and an efficient semiparametric maximum likelihood (ML) estimator based on an i.i.d. assumption. We derive large sample properties for our estimators. Finally, we demonstrate the model fitting performance and its empirical relevance on an application. We investigate how the different components of the illiquidity process obtained from our model relate to the stock market risk premium using data on the S&P 500 stock market index.

Keywords: Kernel; Nonparametric Estimation; Semiparametric Model.

JEL Classification: C12, C14

[*]Email: `christian.hafner@uclouvain.be`.

[†]Email: `obl20@cam.ac.uk`.

[‡]Corresponding author; Email: `lw711@cam.ac.uk`; Address: Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD.

1 Introduction

Liquidity is a fundamental property of a well-functioning market, and lack of liquidity is generally at the heart of many financial crises and disasters. Common ways of measuring liquidity using high-frequency data include bid-ask spreads, effective spreads, realized spreads, depth, weighted depth, and transaction volume. There is a large literature that uses such measures to compare market quality across markets, across time, and before and after interventions of various sorts. For example, it has been a big part of the debate around high-frequency trading, i.e., whether such trading activity has improved or degraded market liquidity, see e.g. Hendershott et al. (2011), O'Hara and Ye (2011). There are many complex issues in working with high-frequency trade and quote data. Instead, there are several methods widely used to measure liquidity using lower frequency data, i.e. daily transaction price data, see Goyenko et al. (2009) for a review. We focus on the Amihud illiquidity measure as proposed in Amihud (2002) which has proven to be very popular in the empirical literature. It is easy to implement and by all accounts relatively robust. It has been shown to influence the cross-sectional asset returns through the so-called illiquidity premium, see Amihud and Mendelson (2015).

We propose a dynamic semiparametric model for illiquidity measured by the daily component of the Amihud measure. Specifically, we propose a multiplicative error model (MEM) that contains a nonparametric long run trend and a parametric short-run autoregressive process as in Engle et al. (2012). The trend part is important for many datasets where liquidity has improved in a secular fashion such as the S&P 500 over the last hundred years and Bitcoin over the much more recent period of its operation. The nonparametric trend is comparable with the conventional monthly averaged measure widely used in the literature, except that our measure is available daily and the implicit length of averaging is controlled by a bandwidth parameter to be chosen by practitioners. Further, the dynamic component measures the short-run variation in liquidity that may be of equal interest.

We approach estimation of the parametric part through GMM based on the first conditional moment restriction, which is consistent under minimal conditions, as well as through a semiparametric likelihood procedure that assumes i.i.d. shocks. In the latter approach, we consider two cases, one where the shock distribution is parametric such as the Weibull or Burr distributions and a further case in which the shock distribution is not specified and is treated nonparametrically. The Burr distribution and the nonparametric distribution allow for heavy tails that we might see during liquidity crises. We develop the distribution theory for the three cases to enable valid inference.

We use the five largest US technology company stocks and the Bitcoin asset to demonstrate the model performance in terms of fitting relevant features of the illiquidity data and we provide various model diagnostics and specification tests. We show that the efficient semiparametric maximum likelihood estimator, assuming a parametric Weibull or Burr distribution for the error term, captures most of the salient features of the illiquidity process. This can be further improved by using a nonparametric density estimator.

We also investigate how the different components of the illiquidity process obtained from our model relate to the stock market risk premium using data on the S&P 500 stock market index. We find that the detrended market risk premium is positively affected by the anticipated short-run illiquidity process and negatively associated with the unanticipated component of market illiquidity. This observation is in agreement with the results of Amihud (2002) which were based on an autoregressive (AR) model fit to monthly illiquidity.

The Multiplicative error model has been applied to many different positive-valued financial time series including volatility, duration between trades, and transaction volume, see e.g. Engle (2002). The MEM model and its applications and developments over the last 20 years are reviewed in Cipollini and Gallo (2022). The VLAB applies this model and provides regular updates on their website (`https://vlab.stern.nyu.edu/liquidity`) for a number of series according to their specific implementation. We give a more detailed comparison of our model with theirs in Appendix A of Hafner et al. (2023).

The remainder of the paper is organized as follows. In Section 2, we discuss the time series properties of the Amihud illiquidity measure and introduce our DArLiQ model. In Section 3, we discuss estimation via GMM based on conditional moment restrictions and through a semiparametric maximum likelihood procedure that assumes i.i.d. shocks. Large sample properties of our procedures are provided in Section 4. We provide a Monte Carlo study in Section 5 to analyze the finite sample performance of the GMM and semiparametric ML estimation procedures. Section 6 presents a detailed empirical application. The appendices are delegated to a separate file which is available online as Hafner et al. (2023).

2 Amihud illiquidity and the model

The Amihud (2002) illiquidity measure of a stock at time t, A_t, is defined as

$$A_t = \frac{1}{n_t} \sum_{j=1}^{n_t} \ell_{t_j}, \quad \ell_{t_j} = \frac{|R_{t_j}|}{V_{t_j}}, \tag{1}$$

where R_{t_j} is the stock return and V_{t_j} is the (dollar) trading volume at time t_j, $j = 1, ..., n_t$. Typically, the measure is computed over monthly or lower frequencies by averaging the daily illiquidity ratio ℓ_{t_j} over n_t observations within the period of interest, e.g. n_t being the number of trading days within a month. Intuitively, the Amihud measure captures the fact that a stock is less liquid if a given trading volume generates a larger move in its price. The Amihud illiquidity measure is a good proxy for high-frequency measures of price impact (Goyenko et al. (2009); Hasbrouck (2009)) with the advantage of only requiring daily price and volume data. Barardehi et al. (2021) proposed to replace the close-to-close return by the overnight component of that return. Fong et al. (2018) proposed a more general class of liquidity measures based on ratios of functions of volatility to functions of trading volume.

Both modifications can easily be accommodated in our framework, but we focus on the original Amihud measure as this is currently the most popular approach.

Empirical evidence points to the existence of factors driving low-frequency variations in illiquidity dynamics in addition to higher-frequency variations. To illustrate, we plot in Figure 1 the evolution of daily log Amihud illiquidity measure for S&P 500 over 1950–2021. We observe that the illiquidity series exhibits a strong downward trend over time, at least up to 2005. Trends in illiquidity series are not limited to S&P 500, as we show in Appendix B using Fab5 and Bitcoin series. This evidence motivates our framework for the Amihud illiquidity measure which weakens the requirement on stationarity. We develop a new class of dynamic autoregressive liquidity (DArLiQ) models, which captures both the slow-varying long-term trend and short-run autoregressive component relevant for illiquidity modelling.

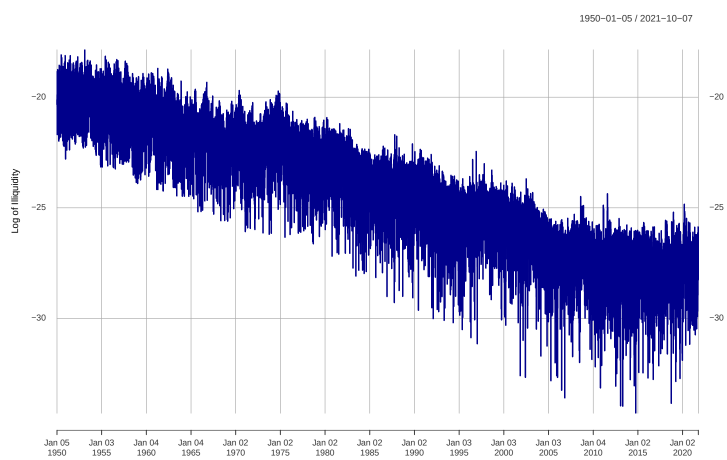

Figure 1: S&P 500 index daily log illiquidity – $\log \ell_t$.

Suppose that ℓ_t is a non-negative process that follows a multiplicative process as in Engle and Gallo (2006) but possesses a nonparametric multiplicative component to account for nonstationarity or trend as in Engle and Rangel (2008) and Hafner and Linton (2010). Let

$$\ell_t = g(t/T)\lambda_t\zeta_t \tag{2}$$

$$\lambda_t = \omega + \beta\lambda_{t-1} + \gamma\ell^*_{t-1}, \tag{3}$$

where $g(.)$ is a positive and smooth but unknown function of rescaled time, $\ell_t^* = \ell_t/g(t/T)$ is the rescaled liquidity, and ζ_t is a non-negative random variable with conditional mean one and finite unconditional variance denoted as σ_ζ^2. We present evidence later that the assumption of finite unconditional variance for the shock process is reasonable. Extensions to higher order models, including more lags of λ_t and ℓ_t^* are straightforward and analogous to the ARCH literature. Note that $\omega > 0,\ \beta \geq 0,\ \gamma \geq 0$ are sufficient conditions for $\lambda_t > 0$ with probability one. Furthermore, provided $\beta+\gamma < 1$, the process ℓ_t^* is stationary in mean (and perforce strictly stationary) and follows an ARMA(1,1) process (with heteroskedastic errors). If $\beta + \gamma \geq 1$, then the process ℓ_t^* is not weakly stationary although it is strongly stationary for a range of such parameter values. More problematic is that in this case $E(\ell_t)$ ceases to exist and our estimation strategy, which is based directly on moment restrictions will fail. This can be addressed using the approach based on quantile restrictions developed as in Koo and Linton (2015). In this paper, we will focus on our mean based approach as evidence show that ℓ_t possesses several moments and so our main strategy is reasonable for most datasets. We further show in Appendix J that the normalized trend functions obtained using the mean and median smoothing methods are comparable and the choice of the smoothing method has a minor impact on the parameter estimates for the λ_t process.

There is an identification issue because we can multiply and divide the two components g, λ by the same constant. We suppose that $E(\lambda_t) = 1$, which is achieved by setting $\omega = 1-\beta-\gamma$. The series $\ell_t^* = \lambda_t\zeta_t$ possesses the same stationarity properties as ℓ_t from the model without a trend. We may suppose that the error process ζ_t is i.i.d. with some c.d.f F. Francq and Zakoïan (2006) (Theorems 2 and 3) ensures that the process ℓ_t^* is strictly stationary and geometrically ergodic under our restrictions on β, γ. The i.i.d. assumption can be helpful for estimation but it may also be important for the calculation of "Liquidity at Risk", which would require further assumptions about the conditional quantiles of ζ_t. Note that the process ℓ_t actually depends on T and forms a triangular array, $\ell_{t,T}$, but for notational economy, we suppress this dependency. The process may be initialized from its stationary distribution or from some fixed values $\{\ell_0, \lambda_0\}$.

3 Estimation

We suppose that a sample of non-negative $\ell_t, t = 1, \ldots, T$ is observed. We will throughout maintain that ℓ_t possesses uniformly bounded moments up to some order, which as acknowledged requires some restriction on the dynamic parameters. Under these conditions, estimation is guided by assumptions made about the error process ζ_t. The minimalist approach is to assume only that with probability one $E(\zeta_t - 1|\mathcal{F}_{t-1}) = 0$, and $E(\zeta_t^2) <= C < \infty$. In that case, one can estimate the function $g(.)$ by conditional mean smoothing of ℓ_t and the identified parameters β, γ by the GMM approach. Provided that additional high level weak dependence conditions are satisfied, one can ensure a CLT for the resulting estimators. As in Cipollini et al. (2013), one may wish to additionally specify a second conditional moment restriction whereby $E(\zeta_t^2 - (1+\sigma_\zeta^2)|\mathcal{F}_{t-1}) = 0$ with probability one. This additional moment restriction permits more efficient estimation provided that this restriction is true, but if it is not true, using this additional moment restriction will bias the parameter estimates.

We may further assume that ζ_t is i.i.d. which implies the conditional moment restriction but also other restrictions. Consider the mixed continuous/discrete case where for all $x \geq 0$,

$$\Pr(\zeta_t \leq x) = \pi \mathbb{1}(x = 0) + (1 - \pi)\mathbb{1}(x > 0) \int_0^x f(u)du, \tag{4}$$

where f is an absolutely continuous density function with support $(0, \infty)$. In some cases, the discrete component (zero returns) is important while in others it is not. For estimation, we may either assume that f is of unknown form or assume that f is parametrically specified, i.e., f_φ for some unknown shape parameter φ such as Weibull or Burr. For forecasting future values of ℓ_t, one does not need the shock distribution, but prediction intervals and LAR (Liquidity at Risk) require the estimation of some features of the error distribution.

3.1 Estimation based on conditional moment restriction

We first convert the conditional moment restriction $E(\ell_t|\mathcal{F}_{t-1}) = g(t/T)\lambda_t$ to the unconditional moment restriction $E(\ell_t) = g(t/T)$, $t = 1, \dots, T$. We use this condition to obtain an initial consistent estimator of g by the kernel smoothing method, specifically we let

$$\widehat{g}(u) = \frac{1}{T}\sum_{t=1}^{T} K_h(t/T - u)\ell_t, \quad u \in (0, 1), \tag{5}$$

where K is a kernel function symmetric about zero supported on $[-1, 1]$ satisfying $\int K(u)du = 1$, while h is a bandwidth. Because of the equally spaced observations in time, the denominator of the Nadaraya-Watson estimator is unnecessary here for interior points. In this context, the kernel estimator is Best Linear Minimax (under i.i.d. errors) at any fixed $u \in (0, 1)$ according to Fan (1993), i.e. it is equivalent to the local linear estimator. The estimator does suffer from boundary bias and in particular $\widehat{g}(0), \widehat{g}(1)$ will not be consistent without modifications. A standard way to correct for boundary bias is to use boundary kernels that adapt to the estimation point as they approach the boundary, see Gasser et al. (1985). An alternative method is local linear kernel regression, which does not require an explicit boundary correction, see Fan and Gijbels (1996). The issue with both methods is that the estimate of $g(u)$ is not guaranteed to be nonnegative everywhere, whereas the simple estimator is non-negative with probability one. In practice, at least for our application, this does not seem to be an issue and we use also the local linear method. Nevertheless, for our theoretical results we work with $\widehat{g}(u)$ as above for all $u \in [h, 1-h]$ and in the boundary region we either renormalize by $\sum_{t=1}^{T} K_h(t/T - u)/T$ (Nadaraya-Watson estimator) or we set $\widehat{g}(u) = \widehat{g}(h)$ if $u \leq h$ and set $\widehat{g}(u) = \widehat{g}(1-h)$ if $u \geq 1-h$. In this case, we guarantee positivity of our estimate but suffer some performance loss at the boundary. Another advantage of the simple estimator is that one can interpret the widely computed measure A_t defined in (1) as a special case of $\widehat{g}(u)$ with uniform kernel and bandwidth equivalent to a month around point u. We define the detrended liquidity $\widehat{\ell}_t^* = \ell_t/\widehat{g}(t/T)$, $t = 1, \dots, T$.

We next estimate the dynamic parameters $\theta = (\beta, \gamma)^\intercal$ by exploiting the conditional moment restriction $E(\ell_t^*|\mathcal{F}_{t-1}) = \lambda_t$, where $\ell_t^* = \ell_t/g(t/T)$, $t = 1, \ldots, T$. In practice, we define for any $\theta \in \Theta$, where Θ is a compact set defined below, $\widehat{\lambda}_t(\theta) = 1-\beta-\gamma+\beta\lambda_{t-1}+\gamma\widehat{\ell}_{t-1}^*$ for $t = 1, \ldots, T$, where we take initializations $\widehat{\ell}_0^*, \lambda_0 = 1$ for simplicity. Then we define $\rho_t(\theta, \widehat{g}) = z_{t-1}(\widehat{\ell}_t^* - \widehat{\lambda}_t(\theta))$, where $z_{t-1} \in \mathcal{F}_{t-1}$ are instruments, and let

$$\widehat{\theta}_{GMM} = \arg\min_{\theta\in\Theta} \left\| M_T(\theta, \widehat{g}) \right\|_W, \quad M_T(\theta, \widehat{g}) = \frac{1}{T}\sum_{t\in I_T} \rho_t(\theta, \widehat{g}),$$

where W is a weighting matrix and $I_T \subset \{1, \ldots, T\}$. In the sequel we suppress the notation I_T, although we discuss this issue in the Supplementary material. The estimator $\widehat{\theta}_{GMM}$ is consistent for θ under our conditions below (and we drop the subscript GMM below). We suggest an improved estimator of $g(.)$. We work from conditional moment restriction $E\left(\frac{\ell_t}{\lambda_t}|\mathcal{F}_{t-1}\right) = g(t/T)$, which is now feasible given our consistent estimates of θ and hence λ_t. We take a simple implementation of local GMM, Gozalo and Linton (2000) and Lewbel (2007), based only on constant instruments in which case we obtain the closed form

$$\widetilde{g}(u) = \frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T - u)\frac{\ell_t}{\widehat{\lambda}_t}, \quad u \in (0, 1), \tag{6}$$

where $\widehat{\lambda}_t = \widehat{\lambda}_t(\widehat{\theta}, \widehat{g})$ are estimated in the previous procedure. The kernel K is as before but the bandwidth sequence $\widetilde{h}$ may be different reflecting the different bias variance trade-off.

3.2 Estimation based on i.i.d. assumption

We assume that the error ζ_t is i.i.d. with mean one, variance σ_ζ^2, and c.d.f. F as specified above. The semiparametric case where the density f is unknown has been treated in other time series models by Kreiss (1987), Linton (1993), Drost and Klaassen (1997), and Ling and McAleer (2003). For a given density f, define the Fisher scale score and information:

$$s_2(\zeta) = -\left(1 + \zeta\frac{f'(\zeta)}{f(\zeta)}\right), \quad I_2(f) = \int s_j^2(\zeta)f(\zeta)d\zeta. \tag{7}$$

3.2.1 Parametric density case

Suppose that ζ_t follows a parametric distribution with density f depending on some unknown parameters φ, denoted as f_φ. Among potential candidates, Weibull and Gamma distributions have proven to be solid choices for duration modelling, e.g. in Engle and Russell (1998), as they are more flexible than the exponential distribution and allow for over- or under-dispersion. In case of fat-tailed data, a popular choice in the duration literature has been the Burr distribution, which nests the Weibull and allows for fat tails, see e.g. Bauwens and Giot (2001). Analyzing the key features of the fitted shock in the GMM case can provide some indications on which distribution to choose. We elaborate on this point in the empirical study.

If $g(.)$ were known, the log likelihood function of $\{\ell_1, \dots, \ell_T\}$ is, apart from a term to do with $g(.)$ which does not depend on parameters, equal to

$$L(\theta, \varphi, \pi|\ell_1, \dots, \ell_T) = \sum_{\ell_t=0} \log \pi + \sum_{\ell_t>0} \log(1-\pi) - \sum_{\ell_t>0} \log \lambda_t(\theta) + \sum_{\ell_t>0} \log f_\varphi\left(\zeta_t(\theta)\right),$$

where $\zeta_t(\theta) = \ell_t/\lambda_t(\theta)g(t/T)$. From this we can see the separability of π; the parameter π may be estimated by simply counting the frequency of zeros of ℓ_t.[1] The remaining quantities are estimated using non-zero observations only. To avoid complicating the notation, we assume in the sequel that $\pi = 0$. In practice, given a consistent estimate of $g(.)$, we may maximize an estimated version of likelihood $\widehat{L}(\theta, \varphi)$, where $g(.)$ is replaced by $\widehat{g}(.)$ or $\widetilde{g}(.)$. In fact, we avoid further nonlinear optimization by using our initial consistent estimates of θ and auxiliary initial consistent estimates of φ, $\widehat{\varphi}$, which may be obtained through closed-form moment estimators. For example, in the Gamma case, parameterized to have mean one, the parameter φ can be estimated as one over the sample variance of the residuals.

We show in Appendix E.2 that the efficient score functions (in the semiparametric

[1] We adopt this approach in Appendix H where we investigate the occurrence of zero returns in S&P 500 over 1950-2021. The data contains 125 zeros (0.69% of the sample) and majority occurred before 2000.

model) for $\eta = (\theta^\intercal, \varphi)^\intercal$ in the presence of unknown $g(.)$ are:

$$L_\theta^*(\eta) = \sum_{t=1}^T \ell_{\theta t}^*(\eta), \quad \ell_{\theta t}^*(\eta) = s_2(\zeta_t)\left(\frac{\partial \log \lambda_t}{\partial \theta} - \frac{E\left[\frac{\partial \log \lambda_t}{\partial \theta}\frac{1}{\lambda_t}\right]}{E\left(\frac{1}{\lambda_t^2}\right)}\frac{1}{\lambda_t}\right), \tag{8}$$

$$L_\varphi^*(\eta) = \sum_{t=1}^T \ell_{\varphi t}^*(\eta), \quad \ell_{\varphi t}^*(\eta) = \frac{\partial \log f_\varphi(\zeta_t)}{\partial \varphi} - \frac{E\left(\frac{\partial \log f_\varphi(\zeta_t)}{\partial \varphi}s_2(\zeta_t)\right)E\left(\frac{1}{\lambda_t}\right)}{I_2(f_\varphi)E\left(\frac{1}{\lambda_t^2}\right)}s_2(\zeta_t)\frac{1}{\lambda_t}. \tag{9}$$

To obtain fully efficient estimates of η, we use one-step updating from initial root-T consistent estimates, Bickel (1982), Bickel et al. (1993), Linton (1993), Drost and Klaassen (1997), and Ling and McAleer (2003). Denote $\widetilde{\eta} = (\widetilde{\theta}^\intercal, \widetilde{\varphi})^\intercal$, $\widehat{\eta} = (\widehat{\theta}^\intercal, \widehat{\varphi})^\intercal$, and let $\ell_{\eta t}^* = (\ell_{\theta t}^{*\intercal}, \ell_{\varphi t}^*)^\intercal$, then let

$$\widetilde{\eta} = \widehat{\eta} + \mathcal{I}_{\eta\eta}^*(\widehat{\eta}, \widehat{g})^{-1} S_\eta^*(\widehat{\eta}, \widehat{g}), \tag{10}$$

$$\mathcal{I}_{\eta\eta}^*(\widehat{\eta}, \widehat{g}) = \frac{1}{T}\sum_{t=1}^T \ell_{\eta t}^*(\widehat{\eta}, \widehat{g})\ell_{\eta t}^*(\widehat{\eta}, \widehat{g})^\intercal, \quad S_\eta^*(\widehat{\eta}, \widehat{g}) = \frac{1}{T}\sum_{t=1}^T \ell_{\eta t}^*(\widehat{\eta}, \widehat{g}),$$

where $\ell_{\theta t}^*(\widehat{\eta}, \widehat{g})$ and $\ell_{\varphi t}^*(\widehat{\eta}, \widehat{g})$ are given by $\ell_{\theta t}^*(\widehat{\eta}, \widehat{g}) = \widehat{s}_2(\widehat{\zeta}_t)\left(\frac{\partial \log \widehat{\lambda}_t}{\partial \theta} - \frac{\frac{1}{T}\sum_{t=1}^T \frac{\partial \log \widehat{\lambda}_t}{\partial \theta}\frac{1}{\widehat{\lambda}_t}}{\frac{1}{T}\sum_{t=1}^T \frac{1}{\widehat{\lambda}_t^2}}\frac{1}{\widehat{\lambda}_t}\right)$ and $\ell_{\varphi t}^*(\widehat{\eta}, \widehat{g}) = \frac{\partial \log f_{\widehat{\varphi}}(\widehat{\zeta}_t)}{\partial \varphi} - \frac{\frac{1}{T}\sum_{t=1}^T \frac{\partial \log f_{\widehat{\varphi}}(\widehat{\zeta}_t)}{\partial \varphi}\widehat{s}_2(\widehat{\zeta}_t)\frac{1}{T}\sum_{t=1}^T \frac{1}{\widehat{\lambda}_t}}{\frac{1}{T}\sum_{t=1}^T \widehat{s}_2^2(\widehat{\zeta}_t)\frac{1}{T}\sum_{t=1}^T \frac{1}{\widehat{\lambda}_t^2}}\widehat{s}_2(\widehat{\zeta}_t)\frac{1}{\widehat{\lambda}_t}$ with $\widehat{\lambda}_t = 1 - \widehat{\beta} - \widehat{\gamma} + \widehat{\beta}\widehat{\lambda}_{t-1} + \widehat{\gamma}\frac{\ell_{t-1}}{\widehat{g}((t-1)/T)}$ and $\widehat{s}_2(\zeta) = -\left(1 + \zeta\frac{f'_{\widehat{\varphi}}(\zeta)}{f_{\widehat{\varphi}}(\zeta)}\right)$.

The i.i.d. structure also permits one to improve the estimation of g by using the local likelihood method of Tibshirani and Hastie (1987). Suppose that f, θ were known, then the local likelihood estimator of $g(u)$ based on data ℓ_t is given by the maximizer of

$$\sum_{t=1}^T K_h(t/T - u)\left(\log f\left(\zeta_t(g)\right) - \log g\right), \quad \zeta_t(g) = \frac{\ell_t}{\lambda_t g}, \quad t = 1, \ldots, T, \tag{11}$$

with respect to the parameter $g \in \mathbb{R}_+$. In general, this involves nonlinear optimization with respect to the scalar parameters. Instead, we will pursue a one-step updating approach. Following Fan and Chen (1999), we may update the estimator of g by

$$\widetilde{g}_{LocL}(u) = \widehat{g}(u) - \widehat{L}_{gg}^{-1}(\widehat{g}(u); u)\widehat{L}_g(\widehat{g}(u); u), \tag{12}$$

where we have $\widehat{L}_g(g; u) = \partial\widehat{L}(g; u)/\partial g$ and $\widehat{L}_{gg}(g; u) = \partial^2\widehat{L}(g; u)/\partial g^2$ with $\widehat{L}(g; u) = \sum_{t=1}^T K_{h^*}(t/T - u)\left(\log f_{\widehat{\varphi}}\left(\widetilde{\zeta}_t(g)\right) - \log g\right)$ and $\widetilde{\zeta}_t(g) = \frac{\ell_t}{g\lambda_t(\widehat{\theta}, \widehat{g})}$, $t = 1, \ldots, T$.

3.2.2 Nonparametric density case

In Appendix E.3, we derive the efficient score function for θ in the semiparametric model with unknown f, g, thereby extending Drost and Werker (2004). This is $L_\theta^{**}(\theta) = \sum_{t=1}^T \ell_{\theta t}^{**}(\theta)$ with $\ell_{\theta t}^{**}(\theta) = \left(\frac{\zeta_t - 1}{\sigma_\zeta^2} + s_2(\zeta_t)\right) a + s_2(\zeta_t)\left(\frac{\partial \log \lambda_t(\theta)}{\partial \theta} - b\frac{1}{\lambda_t}\right) = \frac{\zeta_t - 1}{\sigma_\zeta^2} a + s_2(\zeta_t)\left(\frac{\partial \log \lambda_t(\theta)}{\partial \theta} - a - b\frac{1}{\lambda_t}\right)$ for some a, b with $a = E\left(\frac{\partial \log \lambda_t}{\partial \theta}\right) - bE\left(\frac{1}{\lambda_t}\right)$ and $b = \frac{E\left(\frac{1}{\lambda_t}\frac{\partial \log \lambda_t}{\partial \theta}\right) - \kappa E\left(\frac{\partial \log \lambda_t}{\partial \theta}\right) E\left(\frac{1}{\lambda_t}\right)}{E\left(\frac{1}{\lambda_t^2}\right) - \kappa E^2\left(\frac{1}{\lambda_t}\right)}$ where $\kappa = 1 - 1/I_2(f)\sigma_\zeta^2$.

Suppose we have initial consistent estimators of $\theta, g(.)$. Then, one can estimate the density function $f(\zeta)$ by $\widehat{f}(\zeta) = \frac{1}{T}\sum_{t=1}^T K_{h_f}\left(\widehat{\zeta}_t - \zeta\right)$, where h_f is another bandwidth sequence. The residuals are defined as $\widehat{\zeta}_t = \ell_t/\widehat{g}(t/T)\widehat{\lambda}_t$, $t = 1, \ldots, T$. This estimator does not impose restriction $E(\zeta_t) = 1$. Instead, we also consider estimator based on the rescaled residuals $\widehat{\zeta}_t / \sum_{t=1}^T \widehat{\zeta}_t / T$. Notationally, we assume the same kernel as in the estimation of the liquidity trend but this need not be the case. In particular, since $\zeta_t \geq 0$ one may wish to use special kernel methods adapted to this problem, Chen (2000) and Scaillet (2004).

We propose to construct efficient estimators of θ by two step estimation based on initial consistent estimates of θ, f, g. Specifically, let:

$$\widetilde{\widehat{\theta}} = \widehat{\theta} + \mathcal{I}_{\theta\theta}^{**}(\widehat{\theta}, \widehat{f}, \widehat{g})^{-1} S_\theta^{**}(\widehat{\theta}, \widehat{f}, \widehat{g}), \tag{13}$$

$$\mathcal{I}_{\theta\theta}^{**}(\widehat{\theta}, \widehat{f}, \widehat{g}) = \frac{1}{T}\sum_{t=1}^T \ell_{\theta t}^{**}(\widehat{\theta}, \widehat{f}, \widehat{g}) \ell_{\theta t}^{**}(\widehat{\theta}, \widehat{f}, \widehat{g})^\intercal \widehat{1}_t, \quad S_\theta^{**}(\widehat{\theta}, \widehat{f}, \widehat{g}) = \frac{1}{T}\sum_{t=1}^T \ell_{\theta t}^{**}(\widehat{\theta}, \widehat{f}, \widehat{g}) \widehat{1}_t,$$

where $\ell_{\theta t}^{**}(\widehat{\theta}, \widehat{f}, \widehat{g}) = \frac{\widehat{\zeta}_t - 1}{\widehat{\sigma}_\zeta^2}\widehat{a} + \widehat{s}_2(\widehat{\zeta}_t)\left(\frac{\partial \log \widehat{\lambda}_t}{\partial \theta} - \widehat{a} - \widehat{b}\frac{1}{\widehat{\lambda}_t}\right)$ with $\widehat{\sigma}_\zeta^2 = \sum_{t=1}^T (\widehat{\zeta}_t - \overline{\widehat{\zeta}})^2/T$, $\overline{\widehat{\zeta}} = \sum_{t=1}^T \widehat{\zeta}_t/T$ and $\widehat{\lambda}_t = \lambda_t(\widehat{\theta}, \widehat{g})$. Additionally, we have $\widehat{a} = \frac{1}{T}\sum_{t=1}^T \frac{\partial \log \widehat{\lambda}_t}{\partial \theta} - \widehat{b}\frac{1}{T}\sum_{t=1}^T \frac{1}{\widehat{\lambda}_t}$ and $\widehat{b} = \frac{\frac{1}{T}\sum_{t=1}^T \frac{1}{\widehat{\lambda}_t}\frac{\partial \log \widehat{\lambda}_t}{\partial \theta} - \widehat{\kappa}\frac{1}{T}\sum_{t=1}^T \frac{\partial \log \widehat{\lambda}_t}{\partial \theta}\frac{1}{T}\sum_{t=1}^T \frac{1}{\widehat{\lambda}_t}}{\frac{1}{T}\sum_{t=1}^T \frac{1}{\widehat{\lambda}_t^2} - \widehat{\kappa}\left(\frac{1}{T}\sum_{t=1}^T \frac{1}{\widehat{\lambda}_t}\right)^2}$ where $\widehat{\kappa} = 1 - 1/I_2(\widehat{f})\widehat{\sigma}_\zeta^2$. Here, $\widehat{1}_t$ is a trimming function needed theoretically to reduce the effect of small density estimates. In practice, we have found reasonable performance without trimming. One possible trimming scheme was considered in Linton and Xiao (2007). In the literature on "adaptive estimation", a number of other devices are used primarily to promote simple proofs. These include discretization of the initial estimator and sample splitting, see e.g. Kreiss (1987) and Linton (1993).

We may also update the estimator of g in this case by the one-step improvement

$$\widetilde{\widetilde{g}}_{LocL}(u) = \widehat{g}(u) - \widehat{L}_{gg}^{-1}(\widehat{g}(u); u)\widehat{L}_g(\widehat{g}(u); u),$$

where we have $\widehat{L}_g(g;u) = \partial \widehat{L}(g;u)/\partial g$ and $\widehat{L}_{gg}(g;u) = \partial^2 \widehat{L}(g;u)/\partial g^2$ with $\widehat{L}(g;u) = \sum_{t=1}^{T} K_{h^\dagger}(t/T - u)\left(\log \widehat{f}\left(\widetilde{\zeta}_t(g)\right) - \log g\right)$ and $\widetilde{\zeta}_t(g) = \frac{\ell_t}{g\widehat{\lambda}_t(\widehat{\theta})}$, $t = 1, \ldots, T$, where $h^\dagger$ is another bandwidth sequence.

4 Large sample properties

We suppose that ℓ_t^* is stationary and alpha mixing. This can be shown to hold under the parameter restrictions provided ζ_t is i.i.d. It may also hold when ζ_t itself is only described as a stationary mixing process although this can be difficult to establish; we give further discussion in the Supplementary material. We define the long run variance for a stationary mixing process x_t as $\mathrm{lrvar}(x_t) = \sum_{j=-\infty}^{\infty} \mathrm{cov}(x_t, x_{t-j})$.

Assumption A1. *Suppose that* $g(.) \in \mathcal{G}$, *where for* $c > 0$ $\mathcal{G} = \left\{g : g : [0,1] \to \mathbb{R}_+,\ g(x) \geq c,\ \left|g''(x)\right| < \infty \text{ for all } x \in (0,1), \text{ and } g''_+(0), g''_-(1) \text{ exist}\right\}$. *Define* $||g|| = \left(\int_0^1 g(u)^2 du\right)^{1/2}$ *and* $||g||_\infty = \sup_{u\in[0,1]} |g(u)|$ *for all* $g \in \mathcal{G}$.

Assumption A2. *Suppose that* $\{v_t\}$, *where* $v_t = \lambda_t\zeta_t - 1$, *is a stationary sequence with* $E(v_t) = 0$ *and* $E(|v_t|^{2+\delta}) \leq C < \infty$ *for some* $\delta > 0$. *Furthermore,* v_t *is alpha mixing with for some* $C < \infty$ *and* $\rho > (6 + 2\delta)/\delta$, $\alpha(k) \leq Ck^{-\rho}$.

Assumption A3. *Suppose that* K *is symmetric about zero with compact support* $[-1, 1]$ *such that* $K(\pm 1) = 0$ *and* K *is thrice differentiable where* K''' *is Lipschitz continuous on* $[-1, 1]$. *Let* $||K||^2 = \int_{-1}^1 K(s)^2 ds$, *and* $\mu_j(K) = \int_{-1}^1 s^j K(s) ds$, $j = 0, 1, 2$.

Assumption A4. *Define* $M(\theta, g) = \lim_{T\to\infty} E\left(M_T(\theta, g)\right)$. *For all* $\delta > 0$, *there is an* $\epsilon > 0$ *such that* $\inf_{||\theta-\theta_0||>\delta} \left\|M(\theta, g_0)\right\| \geq \epsilon$. *Uniformly for all* $\theta \in \Theta$, *the function* $M(\theta, g)$

is continuous in g (with respect to the L_2 metric) at $g = g_0$. For all sequences of positive numbers $\delta_T \to 0$, $\sup_{\theta\in\Theta,\|g-g_0\|_\infty\le\delta_T} \left\|M_T(\theta,g) - M(\theta,g_0)\right\| = o_P(1)$.

ASSUMPTION A5. *The ordinary partial derivative and pathwise derivatives* $\Gamma(\theta,g_0) = \frac{\partial M(\theta,g_0)}{\partial\theta}$ and $\Gamma_2(\theta,g_0)\circ(g-g_0) = \frac{\partial}{\partial\tau}M(\theta,g_0+\tau(g-g_0))\Big|_{\tau=0}$ *are assumed to exist in all directions $\theta \in \Theta_\epsilon \subset \Theta, g \in \mathcal{G}_\epsilon \subset \mathcal{G}$, where $\Theta_\epsilon, \mathcal{G}_\epsilon$ are small neighborhoods of θ_0, g_0 respectively. The matrix $\Gamma(\theta,g_0)$ is continuous in θ at $\theta = \theta_0$ and $\Gamma(\theta_0,g_0)$ is of full rank.*

ASSUMPTION A6. *For all positive sequences δ_T, ω_T with $\delta_T \to 0$ and $T^{1/4}\omega_T \to 0$*

(i) $\sup_{\|\theta-\theta_0\|\le\delta_T,\|g-g_0\|\le\omega_T} \omega_T^{-2}\left\|M(\theta,g) - M(\theta,g_0) - \Gamma_2(\theta,g_0)\circ(g-g_0)\right\| \le C$,

(ii) $\sup_{\|\theta-\theta_0\|\le\delta_T,\|g-g_0\|\le\omega_T} \omega_T^{-1}\left\|\Gamma_2(\theta,g_0)\circ(g-g_0) - \Gamma_2(\theta_0,g_0)\circ(g-g_0)\right\| = o(1)$,

(iii) $\sup_{\|\theta-\theta_0\|\le\delta_T,\|g-g_0\|\le\omega_T} \sqrt{T}\left\|M_T(\theta,g) - M(\theta,g) - M_T(\theta_0,g_0)\right\| = o_P(1)$.

Assumption A4 is sufficient for consistency of $\widehat{\theta}$ given that our estimator $\widehat{g}$ is uniformly consistent and $\widehat{g} \in \mathcal{G}$ with probability one. Assumptions A5 and A6 are needed for asymptotic normality of $\widehat{\theta}$. These conditions have been verified in a number of different model settings under more primitive conditions, see Chen et al. (2003). We establish in our treatment of the properties of $\widehat{g}$ that it is uniformly consistent at a rate that can be better than $T^{-1/4}$, which is also required for this theory. The term $\Gamma_2(\theta,g_{x0})\circ(g-g_0)$ determines the correction term in the limiting variance and is established in Appendix D. We also establish in the appendix that $\sqrt{T}\left(M_T(\theta_0,g_0) + \Gamma_2(\theta_0,g_0)\circ(\widehat{g}-g_0)\right)$ satisfies a CLT. This as usual requires undersmoothing of the nonparametric estimation part, e.g. using half of the selected bandwidth. The next two assumptions are needed for the estimators based on the iid error assumption, A7 is for the case where the error density is parametric and A8 is for the case where it is nonparametrically treated.

ASSUMPTION A7. *We suppose that* $\Psi_{k,l}(x;\varphi) = \frac{\partial^{k+l}}{\partial\varphi^k\partial x^l}\log f_\varphi(x)$ *exists and is continuous in both its arguments in a small neighborhood of φ_0 and in all $x \in \mathbb{R}_+$ for*

$k, l = 1, \ldots, 4$ and that $\sup_{\varphi:|\varphi-\varphi_0|\leq cT^{-1/2}} |\Psi_{k,l}(\zeta;\varphi) - \Psi_{k,l}(\zeta;\varphi_0)| \leq R(\zeta)$ for some measurable function $R(.)$, where $E\left(\left(\zeta^l R(\zeta)\right)^\kappa\right), E\left(\left(\zeta^l |\Psi_{k,l}(\zeta;\varphi)|\right)^\kappa\right) < \infty$ for some $\kappa \geq 4$ and for $l = 0, 1$. Furthermore, the efficient information matrix $\mathcal{I}^*_{\eta\eta}(\eta) = \begin{pmatrix} \mathcal{I}^*_{\theta\theta} & \mathcal{I}^*_{\theta\varphi} \\ \mathcal{I}^*_{\varphi\theta} & \mathcal{I}^*_{\varphi\varphi} \end{pmatrix} = \begin{pmatrix} E\left(\ell^*_{\theta t}\ell^{*\intercal}_{\theta t}\right) & E\left(\ell^*_{\theta t}\ell^{*\intercal}_{\varphi t}\right) \\ E\left(\ell^*_{\varphi t}\ell^{*\intercal}_{\theta t}\right) & E\left(\ell^*_{\varphi t}\ell^{*\intercal}_{\varphi t}\right) \end{pmatrix}$ is well defined and positive definite at $\eta = \eta_0$ and continuous in η in a neighborhood of η_0.

ASSUMPTION A8. *We suppose that $f(.)$ is three times continuously differentiable in $x \in \mathbb{R}_+$. Furthermore, the efficient information matrix $\mathcal{I}^{**}_{\theta\theta}(\theta, f, g) = E\left(\ell^{**}_{\theta t}\ell^{**\intercal}_{\theta t}\right)$ is well defined and positive definite at $\theta = \theta_0$ and continuous in θ in a neighborhood of θ_0.*

4.1 Conditional moment restrictions

We first consider the properties of the GMM estimator based on the first conditional moment restriction. This estimator makes the weakest assumptions about the process ζ_t and so it is more robust than the subsequent procedures we analyze.

4.1.1 Nonparametric trend

We first consider the estimator $\widehat{g}(u)$, $u \in (0, 1)$, that is based on smoothing of the raw illiquidity. We may rewrite the model (2), (3) as a nonparametric regression model with trend in mean and variance $\ell_t = g(t/T) + g(t/T)v_t$, where v_t is a mean zero stationary and alpha mixing series. We adapt the results of Francisco-Fernández and Vilar-Fernández (2001) for local polynomial estimators in the case without trending heteroskedasticity to obtain the following central limit theorem.

Theorem 1. *Suppose that A1-A3 hold and $h = cT^{-1/5}$ for $c > 0$. Then for $u \in (0, 1)$*

$$\sqrt{Th}\left(\widehat{g}(u) - g(u) - h^2 b(u)\right) \Longrightarrow N\left(0, V_1(u)\right), \; b(u) = \frac{1}{2}\mu_2(K)g''(u) \; ; \; V_1(u) = g^2(u)||K||^2\sigma^2_{v,\infty}.$$

where $\sigma_{v,\infty}^2$ is the long-run variance of v_t, i.e. $\text{lrvar}(v_t)$. The estimator is consistent and asymptotically normal with optimal rate of $T^{-2/5}$ based on the smoothness assumption. Bandwidth selection and inference procedures require the estimation of $\text{lrvar}(v_t)$, which in general requires further justification. Instead, it may be preferable to work with the refined estimator $\widetilde{g}(u)$ based on the estimator of θ. We have for this estimator the following CLT.

Theorem 2. *Suppose that Assumptions A1-A3 hold, that $\widehat{\theta}$ is $\sqrt{T}$- consistent, that $\widetilde{h} = cT^{-1/5}$ for some $c > 0$ and $Th^5 \to 0$ and $Th/\log T \to \infty$. Then for any $u \in (0,1)$*

$$\sqrt{T\widetilde{h}}\left(\widetilde{g}(u) - g(u) - \widetilde{h}^2 b(u)\right) \Longrightarrow N\left(0, V_2(u)\right), \; b(u) = \frac{1}{2}\mu_2(K)g''(u); \; V_2(u) = g^2(u)||K||^2\sigma_\zeta^2.$$

The bias term is the same as in Theorem 1 by virtue of the undersmoothing of the first step, see Linton and Xiao (2007). The limiting variance is different though and in particular it is proportional to the variance of ζ_t, which is generally smaller and easier to estimate than the long run variance of v_t. When ζ_t is i.i.d., $E(\lambda_t^2(\zeta_t - 1)^2) = E(\lambda_t^2)\sigma_\zeta^2 \geq \sigma_\zeta^2$, because $E(\lambda_t^2) \geq 1$ by the Cauchy-Schwarz inequality since $E(\lambda_t) = 1$. For this estimator, consistent standard errors (assuming undersmoothing) can be based on $\widehat{V}(u) = \widetilde{g}^2(u)||K||^2\widehat{\sigma}_\zeta^2$ where $\widehat{\sigma}_\zeta^2$ is an estimator of σ_ζ^2, e.g. the sample variance of ζ_t after estimation. The omission of the bias effect in the confidence interval has been subject to criticism and debate and many alternative inference approaches have been suggested, at least in the i.i.d. case, see for example Schennach (2015) and Calonico et al. (2018). One simple approach here is to use the pilot model method used in bandwidth selection, Silverman (1986). Specifically, suppose that $g(u) = \exp(a_0 + a_1 u)$ for some unknown parameters a_0, a_1. In that case, $g''(u) = a_1^2 \exp(a_0 + a_1 u)$ and given estimates of a_0, a_1 which can be obtained by the OLS of logarithmic liquidity on a constant and trend (with some adjustment). One can include the estimated bias in the inference procedure.

We may further use this pilot model to select the bandwidth. Since $g''(u)/g(u) = a_1^2$, a "rule-of-thumb" optimal bandwidth procedure would be $h(u) = \left(\frac{||K||^2\widehat{\sigma}^2}{\mu_2^2(K)\widehat{a}_1^4}\right)^{1/5} T^{-1/5}$, where

$\widehat{\sigma}^2 = \widehat{\sigma}^2_{v,\infty}$ if we smooth ℓ_t, e.g. using a Newey-West estimator, and $\widehat{\sigma}^2 = \widehat{\sigma}^2_{\zeta}$ if we work with $\ell_t/\widehat{\lambda}_t$. Here, $h(u)$ happens to be constant across $u \in (0,1)$. One may also consider higher-order polynomials for the log trend to capture different shapes. In this case, the term $\widehat{a}_1^4$ in $h(u)$ should be replaced by the corresponding quantity of $\int_0^1 (g''(u)/g(u))^2 du$. As undersmoothing is required for the first stage bandwidth h in Theorem 2, a possible practical device is to use a fraction of the rule-of-thumb bandwidth, such as a half, which we will employ in our empirical applications.

Hart (1991) showed that conventional cross-validation fails in settings that include our estimator $\widehat{g}(.)$, that is, the usual recipe for selecting h based on minimizing leave-one-out (or equivalently penalized) squared residuals will produce $h \sim 0$. This arises because the serial correlation in error term leads to a bias in the risk estimation. He proposed a modification to address this, that essentially involved estimating the long run variance. We note that our refined estimator $\widetilde{g}(.)$ is not subject to this criticism, since the error term in that case is a martingale difference sequence. This suggests (although we have not proven this here) that standard cross-validation based on $\{\ell_t/\widehat{\lambda}_t\}$ would produce an asymptotically optimal bandwidth choice for $\widetilde{g}(.)$. In summary, the estimator $\widetilde{g}(u)$ has an advantage over $\widehat{g}(u)$ because of the simplicity of handling inference and bandwidth choice. On the other hand, the estimator $\widehat{g}(u)$ is a simple linear estimator and is robust to the specification of λ_t.

4.1.2 Parametric components

Let $\theta = (\beta, \gamma)^\intercal$ be interior values of set Θ, where $\Theta = \{\theta : \tilde{\epsilon} \leq \beta, \gamma, \beta + \gamma \leq 1 - \tilde{\epsilon}\} \subset \mathbb{R}^2$ for some $\tilde{\epsilon} > 0$. This guarantees that for example $\lambda_t^*(\theta) \geq \tilde{\epsilon}$ for all $\theta \in \Theta$. It also ensures that λ_t possesses at least a first moment. In practice, $\tilde{\epsilon}$ is chosen to be machine zero. Define $w_t = \lambda_t(\zeta_t - 1)z_{t-1} + \frac{1-\beta-\gamma}{1-\beta}(\lambda_t\zeta_t - 1)E(z_{t-1})$.

Theorem 3. *Suppose that Assumptions A1-A6 hold, that $\sqrt{T}h^2 \to 0$ and $Th/\log T \to \infty$,* *and that w_t is a stationary mixing process satisfying A2. Then as $T \to \infty$*

$$\sqrt{T}\left(\widehat{\theta}-\theta\right) \Longrightarrow N(0,V),\ V=\left(\Gamma^{\intercal}W\Gamma\right)^{-1}\left(\Gamma^{\intercal}W\Omega W\Gamma\right)\left(\Gamma^{\intercal}W\Gamma\right)^{-1},\ \Omega=\lim_{T\to\infty}\operatorname{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}w_t\right).$$

In general, the asymptotic variance of $\widehat{\theta}$ will depend on the long run variance of process w_t. Inference procedures may be based on the Newey-West method applied to the residuals $\widehat{w}_t=\widehat{\lambda}_t(\widehat{\zeta}_t-1)z_{t-1}+\frac{1-\widehat{\beta}-\widehat{\gamma}}{1-\widehat{\beta}}(\widehat{\lambda}_t\widehat{\zeta}_t-1)\frac{1}{T}\sum_{t=1}^{T}z_t$.

4.2 Restrictions from i.i.d. shocks

We suppose here that ζ_t is i.i.d. with mean one and density f and that we have initial consistent estimators of $g(.),\theta$ available from the GMM procedure described above, say.

In the case where f is parametrically specified with parameters φ, the model is semi-parametric with parameters $\eta=(\theta^{\intercal},\varphi)^{\intercal}$ and unknown function g. We first consider the local likelihood estimator of the trend function based on the estimated $\widehat{\varphi}$.

Theorem 4. *Suppose that assumptions A1-A3 hold and that $\widehat{\eta}$ is $\sqrt{T}$- consistent. Suppose that $h_*=cT^{-1/5}$ for some $c>0$ and that $Th^5\to 0$ and $Th/\log T\to\infty$. Then for any $u\in(0,1)$, the local likelihood estimator satisfies for some bias $b(u)$,*

$$\sqrt{Th_*}\left(\widetilde{g}_{LocL}(u)-g(u)-h_*^2b(u)\right)\Longrightarrow N\left(0,V(u)\right),\ V(u)=||K||^2I_2^{-1}(f)g(u)^2.$$

We note that regarding the estimation of $g(.)$ the asymptotic distribution is the same whether the error density is known or this is estimated parametrically or nonparametrically, Linton and Xiao (2001). This estimator improves on $\widehat{g}(u)$ and $\widetilde{g}(u)$ when i.i.d. assumption is correct by reducing the asymptotic variance, using the standard Cramer-Rao arguments see Tibshirani (1984). We define pointwise confidence bands in Hafner et al. (2023).

We next turn to the properties of the estimated parametric components defined in (10). We need some further regularity conditions, basically smoothness and moment conditions about the parametric density function.

Theorem 5. *Suppose that Assumptions A1-A7 hold and that* $\sqrt{T}h^2 \to 0$ *and* $Th/\log T \to \infty$. *Suppose that* $\widehat{\eta}$ *is* $\sqrt{T}$*-consistent. Then as* $T \to \infty$

$$\sqrt{T}\,(\widetilde{\eta} - \eta) \Longrightarrow N(0, \mathcal{I}^{*}_{\eta\eta}(\eta, g)^{-1}).$$

Furthermore, the asymptotic variance may be estimated consistently by $\mathcal{I}^{*}_{\eta\eta}(\widetilde{\eta}, \widetilde{g})^{-1}$.

In the case where f is nonparametrically specified, the model is semiparametric with parameters θ and unknown functions g, f. We turn to the properties of the estimated parametric components defined in (13).

Theorem 6. *Suppose that Assumptions A1-A6 and A8 hold and that* $\sqrt{T}h^2 \to 0$ *and* $Th/\log T \to \infty$ *and* $\sqrt{T}h_f^2 \to 0$ *and* $Th_f/\log T \to \infty$. *Suppose that* $\widehat{\theta}$ *is* $\sqrt{T}$*-consistent. Then as* $T \to \infty$

$$\sqrt{T}\left(\widetilde{\widetilde{\theta}} - \theta\right) \Longrightarrow N(0, \mathcal{I}^{**}_{\theta\theta}(\theta, f, g)^{-1}).$$

Furthermore, the asymptotic variance may be estimated consistently by $\mathcal{I}^{**}_{\theta\theta}(\widehat{\theta}, \widehat{f}, \widehat{g})^{-1}$.

5 Monte Carlo study

In the Monte Carlo study, we simulate the illiquidity series using the model specified in (2) and (3) with a quadratic polynomial trend, i.e. $g(t/T) = 0.15 - 0.4(t/T) + 0.3(t/T)^2$. In addition, we assume that the error term ζ_t follows a Burr distribution with two shape parameters $\gamma^B = 1.35, \lambda^B = 0.25$. We consider two cases for the λ_t process with $(\beta, \gamma) = (0.85, 0.1), (0.92, 0.07)$. The model is estimated using the GMM approach and the efficient semiparametric ML estimation. For the latter, we examine both the case with a misspecified error distribution (Weibull) and the case with the correctly specified error density (Burr).

We report in Table 1 and Table 2 the estimation accuracy results – measured by bias and standard deviation – for 2000 replications with sample size $T \in \{500, 1000, 2000, 5000, 10000\}$.

Table 1: Bias and standard deviation of parameter estimates: $(\beta, \gamma) = (0.85, 0.1)$.

		GMM		ML-Weibull		ML-Burr	
		β	γ	β	γ	β	γ
Bias	T=500	0.05535	-0.02275	-0.03453	-0.00782	-0.03257	-0.00892
	T=1000	0.04477	-0.01781	-0.02962	-0.00041	-0.02677	-0.00154
	T=2000	0.02837	-0.01410	-0.02507	-0.00125	-0.02278	-0.00157
	T=5000	0.01622	-0.00916	-0.01328	-0.00046	-0.01196	-0.00076
	T=10000	0.00890	-0.00586	-0.00884	-0.00005	-0.00821	-0.00021
StdDev	T=500	0.05315	0.03854	0.08355	0.03566	0.07862	0.03268
	T=1000	0.04244	0.02930	0.04800	0.02357	0.04344	0.02165
	T=2000	0.03826	0.02294	0.03026	0.01643	0.02835	0.01531
	T=5000	0.02673	0.01529	0.01753	0.01023	0.01643	0.00965
	T=10000	0.02076	0.01148	0.01191	0.00705	0.01117	0.00661

Table 2: Bias and standard deviation of parameter estimates where $(\beta, \gamma) = (0.92, 0.07)$.

		GMM		ML-Weibull		ML-Burr	
		β	γ	β	γ	β	γ
Bias	T=500	0.01810	-0.01712	-0.02706	-0.01756	-0.02582	-0.01769
	T=1000	0.01652	-0.01697	-0.02271	-0.00652	-0.02019	-0.00702
	T=2000	0.01152	-0.01500	-0.01697	-0.00529	-0.01527	-0.00528
	T=5000	0.00544	-0.00900	-0.00858	-0.00173	-0.00764	-0.00187
	T=10000	0.00284	-0.00600	-0.00555	-0.00061	-0.00497	-0.00077
StdDev	T=500	0.04327	0.03164	0.08438	0.03064	0.07458	0.02842
	T=1000	0.03590	0.02643	0.03487	0.01774	0.03069	0.01637
	T=2000	0.02868	0.02005	0.01922	0.01186	0.01731	0.01104
	T=5000	0.02111	0.01646	0.00953	0.00720	0.00876	0.00665
	T=10000	0.01459	0.01074	0.00601	0.00479	0.00558	0.00447

Focusing on the bias criterion, we observe that the bias decreases with respect to the sample size. The ML estimates of the β parameter have lower bias than the GMM ones for low levels of persistence ($\beta + \gamma = 0.95$ in Table 1). However, the opposite pattern is observed for higher levels of persistence ($\beta + \gamma = 0.99$ in Table 2). For the γ parameter, the ML method gives more accurate estimates. We further note that the GMM approach tends to overestimate the β parameter while it underestimates the γ parameter, which leads to a more accurate estimation of the overall degree of persistence $\beta + \gamma$ compared to the ML approach. For the standard deviation criterion, we observe that it decreases when the sample size increases and the rate of decrease for the ML estimation method is faster than the GMM one, which roughly corresponds to a factor of $\sqrt{T}$. Moreover, the standard deviation of the ML estimates is overall lower, with the lowest levels achieved by the ML estimation based on the Burr distribution. This observation confirms that the ML estimation with correctly specified distribution provides the most efficient estimates.

6 Empirical study

The ability to accurately model the illiquidity series is useful to quantify conditions in financial markets and track their evolution over time. In our application, we consider the five largest US tech stocks (Fab 5) and Bitcoin to analyze their illiquidity series using our DArLiQ model.[2] We summarize in Table 3 the descriptive statistics of the illiquidity series.[3] The Bitcoin asset is an order of magnitude less liquid compared to the tech stocks during this period and the illiquidity series of Bitcoin is more volatile, exhibits higher skewness and has thicker tails. The five tech companies have comparable levels of liquidity – although Apple stock is slightly more liquid than the others. Moreover, the illiquidity of Amazon stock has higher skewness and thicker tails compared to the other tech companies.

Table 3: Summary statistics for daily illiquidity – $\ell_t \times 10^{10}$.

	Facebook	Amazon	Apple	Google	Microsoft	Bitcoin
Mean	0.0372	0.0313	0.0187	0.0615	0.0424	1.7013
StdDev	0.0295	0.0389	0.0148	0.0499	0.0398	4.1201
Skewness	1.3673	2.5656	1.1146	1.1921	1.6408	4.0626
Kurtosis	6.2978	10.5467	4.1849	4.5345	6.1514	24.9266

[2]We use data retrieved from Yahoo Finance. The sample starts from the first available data point for each asset until October 7th, 2021. Sample R code can be found at `https://github.com/lw1882/DArLiQ`.

[3]To make it comparable, we consider the daily Amihud illiquidity ratios in their common sample period.

We plot in Figures 3 and 4 (Appendix G.1) the illiquidity and log illiquidity series. To manage boundary issues, we use the local linear method to obtain an initial consistent estimator of the trend $g(t/T)$ where we opt for a Gaussian kernel. In this step, we choose the bandwidth according to the rule of thumb derived in Section 4.1.1 and it is reported in the second row of Table 5.[4] The red curves represent the estimated trend functions and their logarithms. Figure 3 shows that the estimated $g(t/T)$ serves as a good approximation for the time-varying mean of the illiquidity series.[5] A strong downward trend is observed for most illiquidity series, indicating an overall improvement in liquidity conditions over time. We notice a temporary worsening in liquidity conditions during significant market events such as the burst of the dot-com bubble and 2007-2009 Global Financial Crisis.

6.1 Estimation results

We introduce the detrended illiquidity series, $\ell_t^* = \ell_t / g(t/T)$, which is assumed to be mean stationary. For the parameters θ of the λ_t process, we first consider a GMM approach, which is based only on the first conditional moment restriction. We then focus on the estimation results using a semiparametric MLE approach based on an i.i.d. shock assumption.

6.1.1 Estimation based on conditional moment restrictions

We use the GMM approach based on the conditional moment restrictions to obtain consistent initial estimates of the λ_t process parameters θ. We consider the minimalist case where the model is estimated using only the first conditional moment restriction, i.e. $E\left[\frac{\ell_t^*}{\lambda_t} - 1 \mid \mathcal{F}_{t-1}\right] = 0$. We further improve the estimates of the $g(t/T)$ function using the estimated $\widehat{\lambda}_t = \widehat{\lambda}_t\left(\widehat{\theta}_{GMM}\right)$ obtained in the previous step. This, in turn, allows us to

[4]We opt for a polynomial of order three for the log trend which allows the trend function to flexibly capture the various shapes exhibited in the data.

[5]Note that the $\log g(t/T)$ is higher than the mean level of $\log \ell_t$ due to Jensen's inequality.

further improve the estimates of the θ parameters. We report the initial and the updated estimates with associated standard errors in Table 4. It can be observed that the parameter estimates are statistically significant. In addition, the sum of the coefficients $\beta + \gamma$ is close to one, indicating high persistence in the short-run dynamics of the illiquidity series.[6]

Table 4: Estimated parameters of the λ_t process based on first moment restriction.

	Facebook	Amazon	Apple	Google	Microsoft	Bitcoin
β	0.967 (0.025)	0.937 (0.017)	0.912 (0.016)	0.978 (0.013)	0.939 (0.015)	0.960 (0.014)
γ	0.027 (0.014)	0.063 (0.017)	0.084 (0.014)	0.022 (0.010)	0.053 (0.010)	0.034 (0.010)
β^U	0.969 (0.022)	0.937 (0.017)	0.914 (0.015)	0.981 (0.011)	0.941 (0.013)	0.960 (0.013)
γ^U	0.027 (0.014)	0.062 (0.016)	0.083 (0.014)	0.019 (0.009)	0.054 (0.010)	0.036 (0.010)
$\widehat{\sigma^U_\zeta}$	0.712	0.717	0.914	0.767	0.724	0.930
tail^U_ζ	7.196	6.428	3.746	5.502	5.997	3.370

Note: The estimated parameters are $\theta = (\beta, \gamma)$ for the λ_t process. The U superscript indicates that the estimates are the ones obtained using the updated trend function. The numbers in parentheses are the standard errors of the corresponding parameter estimates. $\widehat{\sigma^U_\zeta}$ is the sample standard deviation of $\widehat{\zeta}_t$ and tail^U_ζ is the tail index estimated using the regression $\log(\text{Rank} - 1/2) = a - b \log(\text{Size})$ based on the 5% largest fitted shocks $\widehat{\zeta}_t$.

We improve the estimates of the trend function based on the estimated $\widehat{\lambda}_t$ process, i.e. we estimate the $g(.)$ function based on $\ell_t/\widehat{\lambda}_t$. In this step, we choose the bandwidth using a leave-one-out cross validation approach, see e.g. Chu and Marron (1991) for more details. We can observe from Table 5 that the bandwidth to update the trend function is much smaller than the bandwidth used in the initial step as there is less variation in the $\ell_t/\widehat{\lambda}_t$ series.[7] We further plot the log transforms of the initial and updated estimates of the trend function, i.e. $\log g(t/T)$, together with the log illiquidity series in Figure 5. We observe

[6]We note that, in most cases, 1 falls inside the confidence interval of $\beta + \gamma$. This corresponds to the case where the λ_t process is strongly stationary but not weakly stationary. The moment-based estimator of $g(.)$ is not appropriate in this setting, but this can be addressed using the median smoothing method proposed by Koo and Linton (2015). We show in Appendix J that the normalized trend functions obtained with the mean and median smoothing methods are comparable and the choice of the smoothing method has a minor impact on the parameter estimates for the λ_t process.

[7]The values reported in Table 5 are the optimal bandwidth selected according to the rule-of-thumb (h_0) and cross validation approach (h). When undersmoothing is required, we use half of the selected

that the updated trend function estimates are different from the initial estimates but only to a minor extent. In addition, the updated estimates of the λ_t process parameters are slightly different from the initial estimates with the overall degree of persistence $(\beta + \gamma)$ being almost the same. This observation indicates that a 2-step approach consisting in first using a local linear estimator for the trend function and then estimating the λ_t process and its associated parameters θ can be a viable option in empirical applications.

Table 5: Bandwidth choice for the initial estimate and the updated trend function.

	Facebook	Amazon	Apple	Google	Microsoft	Bitcoin
T	2362	6140	10292	4314	8966	2574
h_0	0.038	0.059	0.046	0.092	0.018	0.035
h	0.020	0.018	0.013	0.022	0.013	0.027

Note: T is the total number of observations in the sample period. h_0 is the bandwidth used for the initial estimate of the trend function which is obtained following the rule of thumb. h is the bandwidth used when updating the trend function and is selected using the leave-one-out cross validation approach.

In addition, we compute the sample standard deviation $\widehat{\sigma_\zeta^U}$ of fitted shock series $\widehat{\zeta}_t$ obtained using the updated λ_t parameters. We report the values in the last part of Table 4 and we observe that the sample standard deviations of all assets are below one, indicating under-dispersion. Lastly, we estimate the tail index of the shock series which is between three and four for Apple and Bitcoin while it is between six and eight for the other assets. Our results suggest that the shocks of Apple and Bitcoin have thicker tails.[8]

6.1.2 Estimation: i.i.d. error term with parametric density

We estimate the model using the semiparametric MLE approach where we assume an i.i.d. error term. The conditional distribution of the error term ζ_t can be chosen within the class of distributions satisfying the desired requirements, i.e. the density having non-negative

bandwidth, i.e. $h_0/2$ and $h/2$.

[8]See Appendix I.1 for more details on the tail index analysis.

Table 6: Fully efficient estimates of the λ_t process parameters in the parametric (Weibull) case.

	Facebook	Amazon	Apple	Google	Microsoft	Bitcoin
β	0.885 (0.044)	0.914 (0.003)	0.901 (0.002)	0.927 (0.005)	0.917 (0.004)	0.896 (0.016)
γ	0.058 (0.007)	0.086 (0.003)	0.095 (0.002)	0.061 (0.003)	0.068 (0.002)	0.066 (0.006)
φ	1.366 (0.026)	1.358 (0.013)	1.308 (0.001)	1.266 (0.013)	1.379 (0.011)	1.136 (0.010)
σ_ζ	0.741	0.744	0.771	0.796	0.734	0.882

Note: Estimated parameters are $\theta = (\beta, \gamma, \varphi)$ for λ_t process. φ is the shape parameter of the Weibull distribution with mean 1 and standard deviation σ_ζ of $\sqrt{\frac{\Gamma\left(1+\frac{2}{\varphi}\right)}{\left(\Gamma^2\left(1+\frac{1}{\varphi}\right)\right)} - 1}$. The numbers in parentheses are standard errors.

support with unit mean and finite variance σ_ζ^2. We present estimation results assuming that ζ_t follows a Weibull$(\Gamma(1+\varphi)^{-1}, \varphi)$ distribution with shape parameter φ. Based on the local linear estimator of $g(t/T)$, we first obtain a consistent estimator of the λ_t process parameters via the Quasi-Maximum Likelihood (QML) estimation approach. We then obtain the fully efficient estimates with a one-step update using the efficient scores based on the initial consistent estimators as introduced in Section 3.2.1.

We report the estimated parameters with their standard errors in Table 6. The estimates of the dynamic parameters of λ_t, β and γ, are quite similar across assets, and all illiquidity series exhibit a high degree of persistence as the sum of the estimated coefficients $\beta + \gamma$ is close to one. These findings are quite analogous to univariate GARCH models for volatility. The estimated shape parameters of the Weibull error terms have a higher dispersion across assets and range from 1.14 to 1.38, indicating that the volatility of ζ_t ranges from 0.73 to 0.88. This is consistent with the observation that the five tech stocks have comparable volatilities while Bitcoin has much higher volatility.

We provide diagnostics on the validity of our assumptions on the error term ζ_t. Concerning the i.i.d. assumption, we plot in Figures 6 and 7 (Appendix G.3) respectively the ACF of ζ_t and ζ_t^2. We observe that overall there is no evidence suggesting autocorrelation in the residuals or squared residuals. Moreover, we use the Probability Integral Transform (PIT) to check how well the assumed Weibull distribution fits the data. The histograms of the PITs (shown in Figure 8 Appendix G.3) are quite close to a uniform distribution.

The tail index analysis of the fitted shock series $\widehat{\zeta}_t$ in Section 6.1.1 suggests that the shocks have a thicker tail than the Weibull while exhibiting under-dispersion features. We thus also consider fat-tailed distributions in our analysis, such as the Burr – which nests the Weibull and Lomax – and Inverse Burr distributions. The estimation results in Appendix I.2 show that the Lomax distribution lacks the ability to capture the under-dispersion feature with unit mean restriction. The Burr distribution reduces to the Weibull except for Apple and Bitcoin whose shock terms have thicker tails than the other stocks. The Inverse Burr outperforms the Weibull and Burr in terms of log likelihood except for Microsoft. No distribution among the ones considered above consistently provides a better fit and we thus refrain from searching for more general distributions. Instead, we focus in the next section on whether a more flexible nonparametric density can provide a better fit to the data.

We further improve the estimation of $g(t/T)$ by maximizing the local likelihood based on the estimated $\widehat{\lambda}_t$ and the error density. The log transforms of the initial and updated estimates of the trend function, i.e. $\log g(t/T)$, are plotted in Figure 9 (Appendix G.3) together with the log illiquidity series. As in the GMM case (Appendix G.2), the updated trend function estimates are different from the initial estimate but only to a minor extent.

6.1.3 Estimation: i.i.d. error term with nonparametric density

Table 7: Fully efficient estimates of the λ_t process parameters in the nonparametric case.

	Facebook	Amazon	Apple	Google	Microsoft	Bitcoin
β	0.886 (0.031)	0.914 (0.002)	0.911 (0.002)	0.931 (0.004)	0.921 (0.003)	0.906 (0.010)
γ	0.059 (0.006)	0.085 (0.002)	0.086 (0.002)	0.058 (0.002)	0.067 (0.002)	0.068 (0.005)

Note: The estimated parameters are $\theta = (\beta, \gamma)$ for λ_t process. The numbers in parentheses are standard errors.

We investigate whether replacing the parametric error density f with a nonparametric kernel estimator can further improve the model fit to data. We plot, in Figure 10 (Appendix

G.4), the estimated nonparametric density (using residuals $\widehat{\zeta}_t$ from the GMM case) against the Weibull density using the shape parameter estimate from Section 6.1.2. We observe that the estimated Weibull density curves do not fall between the pointwise two standard deviation bands of the estimated nonparametric densities, suggesting that the difference between the estimated parametric and nonparametric densities is statistically significant.

Table 8: Log likelihood comparison between parametric (Weibull) and nonparametric cases.

	Facebook	Amazon	Apple	Google	Microsoft	Bitcoin
Weibull	-2147.32	-3659.67	-6839.50	-3862.46	-7790.52	-2337.09
Nonparametric	-2111.33	-3594.10	-6618.35	-3800.48	-7767.02	-2315.68
Difference	35.99	65.57	221.15	61.98	23.50	21.41

Note: The difference is $\log LL$ in the nonparametric case minus $\log LL$ in the parametric case.

The estimated nonparametric density allows us to further improve the ML estimation for the λ_t process. We obtain the fully efficient estimates in the nonparametric case using the one-step update approach based on the efficient scores introduced in Section 3.2.2. The estimates and standard errors are reported in Table 7. Comparing the estimated λ_t parameters in the parametric (Table 6) and nonparametric case (Table 7), we observe that the differences are quite small. This indicates that gains in efficiency with respect to GMM, using a one-step update based on the efficient scores, are quite robust with respect to the distribution of ζ_t. However, goodness-of-fit of the parametric and nonparametric models, measured in terms of log-likelihood, is quite different, see results in Table 8. We conclude that the ML estimation assuming a Weibull density for ζ_t provides good performance, but using a nonparametric estimator can further improve the fit in terms of likelihood. In particular, if distributional aspects of the fitted model such as tail properties are important, e.g. for risk management purposes, then the nonparametric model is clearly be preferred.

6.2 Risk premium

Amihud (2002) considers an autoregressive model for annual and monthly market illiquidity and then relates this to the market risk premium. Specifically, he estimates the regression $R_{mt} - R_{ft} = a + b \operatorname{liq}_t^{\mathrm{e}} + c \operatorname{liq}_t^{\mathrm{u}} + \varepsilon_t$, where R_{mt} and R_{ft} are the (annual or monthly) market return and risk-free rate respectively. $\operatorname{liq}_t^{\mathrm{e}}, \operatorname{liq}_t^{\mathrm{u}}$ are the expected and unexpected components of illiquidity determined from the first order autoregression $\operatorname{liq}_t = c_0 + c_1 \operatorname{liq}_{t-1} + \eta_t$, where liq_t is the (annual or monthly) average illiquidity that we called A_t or rather its logarithm. His estimation results strongly support his prior hypothesis that stock excess return is an increasing function of expected illiquidity, and unexpected illiquidity has a negative effect on contemporaneous unexpected stock return (i.e., $b > 0$ and $c < 0$).

We reproduce his analysis within our daily model framework, which has three components to illiquidity: the slowly varying trend, the short run anticipated dynamic component and the unanticipated shock. We first consider the specification for daily stock returns

$$R_{mt} - R_{ft} = a + bg(t/T) + c\lambda_t + d\zeta_t + \varepsilon_t, \tag{14}$$

where $g(.), \lambda_t$ and ζ_t are defined above. This allows the risk premium to depend on all three of the components of illiquidity of our model, see Escanciano et al. (2017) for related specifications. It is difficult to tie together our model for ℓ_t with these regressions (but the same comment would apply to the original work of Amihud (2002)). One criticism might be that we have used returns to construct liquidity and therefore this variable appears on both sides of the regression. However, we can think of returns as being composed of a direction and a magnitude component from the decomposition $R = |R| \operatorname{sign}(R)$. Our liquidity model is about the magnitude $|R|$, (directional information is not used at all by our liquidity model), whereas the dependent variable of regression (14) reflects the direction.

We note that there exists a strong downward trend in illiquidity while returns are somewhat stationary. This suggests that the relationship between the long-run trend of liquidity and the stock excess return would be less significant.[9] Therefore, in the application, we focus on the alternative regression for the detrended equity premium, that is,

$$R_{mt} - R_{ft} - E\left(R_{mt} - R_{ft}\right) = \alpha + \gamma\lambda_t + \delta\zeta_t + \varepsilon_t,$$

To be consistent with our assumptions about λ_t, ζ_t, we should have $\alpha = -(\gamma + \delta)$ but we don't impose this in the estimation although data suggest that $\widehat{\alpha} \simeq -(\widehat{\gamma} + \widehat{\delta})$. In any case, there is a generated regressor issue here when we replace λ_t and ζ_t by their estimated quantities. Therefore, we compute the standard errors for the coefficient estimates using the nonparametric bootstrap procedure outlined in Appendix F.

Table 9: Coefficient estimates of the risk premium regression.

α	γ	δ
-0.076	0.172^{***}	-0.096^{***}
(0.058)	(0.056)	(0.021)

Note: We estimate the regression based on Equation (6.2). Returns are expressed in percentage points. Standard errors are based on the nonparametric bootstrap procedure with 500 replications (see Appendix F for details). Significance level $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

We use the daily historical data of the S&P 500 index to carry out the analysis. The illiquidity and log illiquidity together with the return series are plotted in Figure 11 (Appendix G.5). We report in Table 9 the estimation results based on Equation (6.2).[10] We observe that the estimated γ coefficient for the short-run expected illiquidity component λ_t is positive and significant indicating that the expected market excess return is an increasing function of the short-run expected illiquidity. This is consistent with the intuition that higher expected market illiquidity would make investors demand higher excess returns to compensate for this risk exposure. Moreover, the estimated δ for the shock term ζ_t is

[9]This is confirmed by regression results based on Equation (14). The coefficient estimates for the parameter b associated with the long-run trend illiquidity component are not significant.

[10]The time-varying unconditional equity premium is obtained via a local linear estimator.

negative and significant, suggesting that unexpected market illiquidity has a negative effect on stock excess return. This could be because stock prices would likely fall when illiquidity unexpectedly rises, thus decreasing expected returns.

Acknowledgements

We would like to thank the editor, associate editor and two referees for their insightful comments. We are grateful for discussions with seminar participants at NYU Abu Dhabi, University of York, Humboldt-Universität zu Berlin, Janeway Institute Cambridge and the participants at the 2022 QFFE in Marseille, the conference in honor of Torben G. Andersen in Copenhagen and the 2023 SoFiE conference in Seoul. We would also like to thank Andrew Harvey for his helpful comments. The first author acknowledges the financial support of the ARC research contract 18/23-089 of the Belgian Federal Science Policy Office. The second and third authors acknowledge financial support from the Janeway Institute. The authors report there are no competing interests to declare.

References

Amihud, Y. (2002): "Illiquidity and stock returns: Cross-section and time-series effects". *Journal of Financial Markets* 5.1, pp. 31–56.

Amihud, Y. and Mendelson, H. (2015): "The pricing of illiquidity as a characteristic and as risk". *Multinational Finance Journal* 19.3, pp. 149–168.

Barardehi, Y. H., Bernhardt, D., Ruchti, T. G. and Weidenmier, M. (2021): "The night and day of Amihud's (2002) liquidity measure". *The Review of Asset Pricing Studies* 11.2, pp. 269–308.

Bauwens, L. and Giot, P. (2001): *Econometric Modelling of Stock Market Intraday Activity*. Springer.

BICKEL, P. J. (1982): "On adaptive estimation". *The Annals of Statistics* 10.3, pp. 647–671.

BICKEL, P. J., KLAASSEN, C. A., RITOV, Y. AND WELLNER, J. A. (1993): *Efficient and adaptive estimation for semiparametric models*. Vol. 4. Springer.

CALONICO, S., CATTANEO, M. D. AND FARRELL, M. H. (2018): "On the effect of bias estimation on coverage accuracy in nonparametric inference". *Journal of the American Statistical Association* 113.522, pp. 767–779.

CHEN, S. X. (2000): "Probability density function estimation using gamma kernels". *Annals of the Institute of Statistical Mathematics* 52.3, pp. 471–480.

CHEN, X., LINTON, O. B. AND VAN KEILEGOM, I. (2003): "Estimation of semiparametric models when the criterion function is not smooth". *Econometrica* 71.5, pp. 1591–1608.

CHU, C.-K. AND MARRON, J. S. (1991): "Comparison of two bandwidth selectors with dependent errors". *The Annals of Statistics* 19.4, pp. 1906–1918.

CIPOLLINI, F., ENGLE, R. F. AND GALLO, G. M. (2013): "Semiparametric vector MEM". *Journal of Applied Econometrics* 28.7, pp. 1067–1086.

CIPOLLINI, F. AND GALLO, G. M. (2022): "Multiplicative Error Models: 20 years on". *Econometrics and Statistics*, forthcoming.

DROST, F. C. AND KLAASSEN, C. A. (1997): "Efficient estimation in semiparametric GARCH models". *Journal of Econometrics* 81.1, pp. 193–221.

DROST, F. C. AND WERKER, B. J. M. (2004): "Semiparametric duration models". *Journal of Business & Economic Statistics* 22.1, pp. 40–50.

ENGLE, R. (2002): "New frontiers for ARCH models". *Journal of Applied Econometrics* 17.5, pp. 425–446.

ENGLE, R. F. AND GALLO, G. M. (2006): "A multiple indicators model for volatility using intra-daily data". *Journal of Econometrics* 131.1-2, pp. 3–27.

Engle, R. F., Gallo, G. M. and Velucchi, M. (2012): “Volatility spillovers in East Asian financial markets: A MEM-based approach”. *The Review of Economics and Statistics* 94.1, pp. 222–223.

Engle, R. F. and Rangel, J. G. (2008): “The spline-GARCH model for low-frequency volatility and its global macroeconomic causes”. *The Review of Financial Studies* 21.3, pp. 1187–1222.

Engle, R. F. and Russell, J. R. (1998): “Autoregressive conditional duration: A new model for irregularly spaced transaction data”. *Econometrica* 66.5, pp. 1127–1162.

Escanciano, J. C., Pardo-Fernández, J. C. and Van Keilegom, I. (2017): “Semi-parametric estimation of risk-return relationships”. *Journal of Business & Economic Statistics* 35.1, pp. 40–52.

Fan, J and Gijbels, I. (1996): *Local polynomial modelling and its applications*. Chapman and Hall.

Fan, J. (1993): “Local linear regression smoothers and their minimax efficiencies”. *The Annals of Statistics* 21.1, pp. 196–216.

Fan, J. and Chen, J. (1999): “One-step local quasi-likelihood estimation”. *Journal of the Royal Statistical Society: Series B* 61.4, pp. 927–943.

Fong, K. Y., Holden, C. W. and Tobek, O. (2018): “Are volatility over volume liquidity proxies useful for global or US research?” *Kelley School of Business Research Paper* 17-49.

Francisco-Fernández, M. and Vilar-Fernández, J. M. (2001): “Local polynomial regression estimation with correlated errors”. *Communications in Statistics-Theory and Methods* 30.7, pp. 1271–1293.

Francq, C. and Zakoïan, J.-M. (2006): “Mixing properties of a general class of GARCH (1, 1) models without moment assumptions on the observed process”. *Econometric Theory* 22.5, pp. 815–834.

Gasser, T., Muller, H.-G. and Mammitzsch, V. (1985): "Kernels for nonparametric curve estimation". *Journal of the Royal Statistical Society: Series B* 47.2, pp. 238–252.

Goyenko, R. Y., Holden, C. W. and Trzcinka, C. A. (2009): "Do liquidity measures measure liquidity?" *Journal of Financial Economics* 92.2, pp. 153–181.

Gozalo, P. and Linton, O. B. (2000): "Local nonlinear least squares: Using parametric information in nonparametric regression". *Journal of Econometrics* 99.1, pp. 63–106.

Hafner, C. M. and Linton, O. B. (2010): "Efficient estimation of a multivariate multiplicative volatility model". *Journal of Econometrics* 159.1, pp. 55–73.

Hafner, C. M., Linton, O. B. and Wang, L. (2023): "Supplementary Material for 'Dynamic Autoregressive Liquidity (DArLiQ)'".

Hart, J. D. (1991): "Kernel regression estimation with time series errors". *Journal of the Royal Statistical Society: Series B (Methodological)* 53.1, pp. 173–187.

Hasbrouck, J. (2009): "Trading costs and returns for US equities: Estimating effective costs from daily data". *The Journal of Finance* 64.3, pp. 1445–1477.

Hendershott, T., Jones, C. M. and Menkveld, A. J. (2011): "Does algorithmic trading improve liquidity?" *The Journal of Finance* 66.1, pp. 1–33.

Koo, B. and Linton, O. B. (2015): "Let's get lade: Robust estimation of semiparametric multiplicative volatility models". *Econometric Theory* 31.4, pp. 671–702.

Kreiss, J.-P. (1987): "On adaptive estimation in stationary ARMA processes". *The Annals of Statistics* 15.1, pp. 112–133.

Lewbel, A. (2007): "A local generalized method of moments estimator". *Economics Letters* 94.1, pp. 124–128.

Ling, S. and McAleer, M. (2003): "Asymptotic theory for a vector ARMA-GARCH model". *Econometric theory* 19.2, pp. 280–310.

Linton, O. B. (1993): "Adaptive estimation in ARCH models". *Econometric Theory* 9.4, pp. 539–569.

Linton, O. B. and Xiao, Z. (2001): “Second-order approximation for adaptive regression estimators”. *Econometric Theory* 17.5, pp. 984–1024.

Linton, O. B. and Xiao, Z. (2007): “A nonparametric regression estimator that adapts to error distribution of unknown form”. *Econometric Theory* 23.3, pp. 371–413.

O’Hara, M. and Ye, M. (2011): “Is market fragmentation harming market quality?” *Journal of Financial Economics* 100.3, pp. 459–474.

Scaillet, O. (2004): “Density estimation using inverse and reciprocal inverse Gaussian kernels”. *Nonparametric Statistics* 16.1-2, pp. 217–226.

Schennach, S. M. (2015): “A bias bound approach to nonparametric inference”. *CEMMAP Working Paper* CWP71/15.

Silverman, B. (1986): *Density estimation for statistics and data analysis.* Chapman and Hall, London.

Tibshirani, R. (1984): “Local likelihood estimation”. *PhD Thesis, Stanford University.*

Tibshirani, R. and Hastie, T. (1987): “Local likelihood estimation”. *Journal of the American Statistical Association* 82.398, pp. 559–567.