

Dynamic Autoregressive Liquidity (DARLiQ)

Christian M. Hafner*

Oliver B. Linton[†]

Linqi Wang[‡]

UCLouvain

University of Cambridge

UCLouvain

February 25, 2022

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 GMM estimator based on conditional moment restrictions and an efficient semiparametric ML estimator based on an i.i.d. assumption. We derive large sample properties for both estimators. We further develop a methodology to detect the occurrence of permanent and transitory breaks in the illiquidity process. Finally, we demonstrate the model performance and its empirical relevance on two applications. First, we study the impact of stock splits on the illiquidity dynamics of the five largest US technology company stocks. Second, 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: Nonparametric; Semiparametric; Splits; Structural Change.

JEL Classification: C12, C14

*Louvain Institute of Data Analysis and Modelling in economics and statistics (LIDAM), and ISBA, UCLouvain. Email: christian.hafner@uclouvain.be.

[†]Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD. Email: obl20@cam.ac.uk. Thanks to the Janeway Institute for financial support.

[‡]Louvain Institute of Data Analysis and Modelling in economics and statistics (LIDAM), and LFIN, UCLouvain. Email: linqi.wang@uclouvain.be.

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 big 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. [Brogaard \(2010\)](#), [Hendershott et al. \(2011\)](#), [Beddington et al. \(2012\)](#), [O'Hara and Ye \(2011\)](#). There are many complex issues in working with high frequency trade and quote data in a legally integrated market such as the US, where separate venues exist without synchronized timestamps so that for example establishing the time priority of messages across different venues is difficult. There are several methods widely used to measure liquidity using lower frequency data, i.e., daily data, see [Goyenko et al. \(2009\)](#) for a review of such measures. We focus on the Amihud illiquidity measure as proposed in [Amihud \(2002\)](#). This measure 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 the review of [Amihud and Mendelson \(2015\)](#).

We propose a dynamic semiparametric model for illiquidity as measured by the daily component of the Amihud measure. Specifically, we propose a multiplicative model 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&P500 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 that is 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 the practitioner. Further,

the dynamic component of the model measures the short-run variation in liquidity that may be of equal interest.

We approach estimation through GMM based on the first conditional moment restriction, 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 distribution and a further case in which the shock distribution is not specified and is treated nonparametrically. We develop the distribution theory and efficiency bound in both cases.

We also develop methodology for detecting permanent and transitory changes in liquidity that might arise from structural changes in financial markets such as the upgrade of a stock exchange's matching engine or from stock specific events such as stock splits. In our approach, permanent effects are captured by discontinuous changes in the nonparametric trend function, whereas temporary effects are measured by dummy variables in the dynamic part of the process. We develop the inference tools required to test for the null hypothesis of no changes versus the alternatives of permanent or temporary shifts in the illiquidity process.

In the spirit of [Amihud \(2002\)](#), who studies the effect of expected and unexpected illiquidity on stock excess returns, we also consider the regression modelling of the market risk premium driven by the separate components of liquidity from our model. In particular, we study the link between the stock excess returns and the long-run trend, short-run dynamics as well as the unexpected shocks of the illiquidity process.

We implement our framework on a panel composed by the five largest US technology company stocks and the Bitcoin asset. We demonstrate the model performance in terms of fitting the relevant features of the illiquidity data, and provide various model diagnostics and specification tests. We show that the efficient semiparametric maximum likelihood estimator, assuming a parametric Weibull distribution for the error term, captures well the salient features of the illiquidity process. In addition, we also demonstrate that using a nonparametric density estimator for the error term can further improve the model estimation in terms of likelihood.

We study the impact of stock splits on the illiquidity dynamics of the five largest US technology company stocks. One explanation for why companies split their stock is the theory that this creates “wider” markets, that is, reducing the price level makes it easier for a wider pool of retail investors to buy into the stock and allows existing investors more easily to sell part of their holding to other investors thereby increasing the investor base and the volume of transactions. This in turn should lead to greater liquidity as measured for example by the Amihud measure. However, there are other theoretical arguments presented in [Copeland \(1979\)](#) that may point to a decrease in liquidity following a stock split, and as he says “liquidity changes following stock splits is an empirical question”. [Copeland \(1979\)](#) found: nonstationarities in trading behavior, volume increases less than proportionately, brokerage revenues increased, and increases in proportional bid-ask spreads following stock splits. He argues that “these results lead to the conclusion that there is a permanent decrease in liquidity following the split”. Our results broadly support these findings in our more recent sample data on a special subset of stocks, the tech stocks. Specifically, we document that stock splits cause significant shifts in the long-term illiquidity trend while no significant effects on short-run liquidity dynamics are detected.

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 in agreement with the results of [Amihud \(2002\)](#) (which were based on an AR model fit to monthly illiquidity).

The remainder of the paper is organized as follows. In [Section 2](#), we discuss the Amihud illiquidity measure and its time series properties. [Section 3](#) introduces our DArLiQ model and we discuss in [Section 4](#) estimation via GMM based on the first conditional moment restriction, as well as through a semiparametric likelihood procedure that assumes i.i.d. shocks. The large sample properties of our procedures are provided in [Section 5](#). We develop in [Section 6](#) the methodology to detect permanent and temporary changes in

the liquidity process and in Section 7 the framework to study the effect of illiquidity components on risk premium. Section 8 presents a detailed empirical application of the model, and Section 9 concludes. Theoretical materials including proofs of the theorems are collected in Appendix A to Appendix D. Additional tables and figures for the empirical application are presented in Appendix E.

2 Amihud illiquidity

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

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

where R_t is the stock return and V_t is the (dollar) trading volume at time t . 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. Typically, the measure is computed over periods ranging from a day to a year by averaging the daily illiquidity ratio ℓ_{t_j} over the corresponding period n_t . 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 data on stock prices and trading volumes.

We show in Figure 1 the daily stock log illiquidity series for the five largest US information technology companies (the “Fab 5”) – Amazon, Apple, Facebook, Google, and Microsoft – over the period from May 2012 to October 2021. Note that there is spike in the illiquidity series for Google around end-March 2014 which is caused by a stock split on March 27, 2014.¹ As this event caused irregularity in the trading activities for a few days, we replace the volume data on those dates using the average volume level of the day before and the day after that period. The daily log illiquidity series using the adjusted data are shown in Figure 1b. The illiquidity time series appear broadly stationary during this period, although a slight downward trend can be observed. To gain more insights into

¹The two-for-one stock split was associated with the introduction of a new non-voting share class (Class C shares). See press release.

the conditional dynamics the data, we fit an AR(5) model with a quadratic polynomial trend function to the scaled illiquidity series, $y_t = \ell_t \times 10^{10}$, i.e.

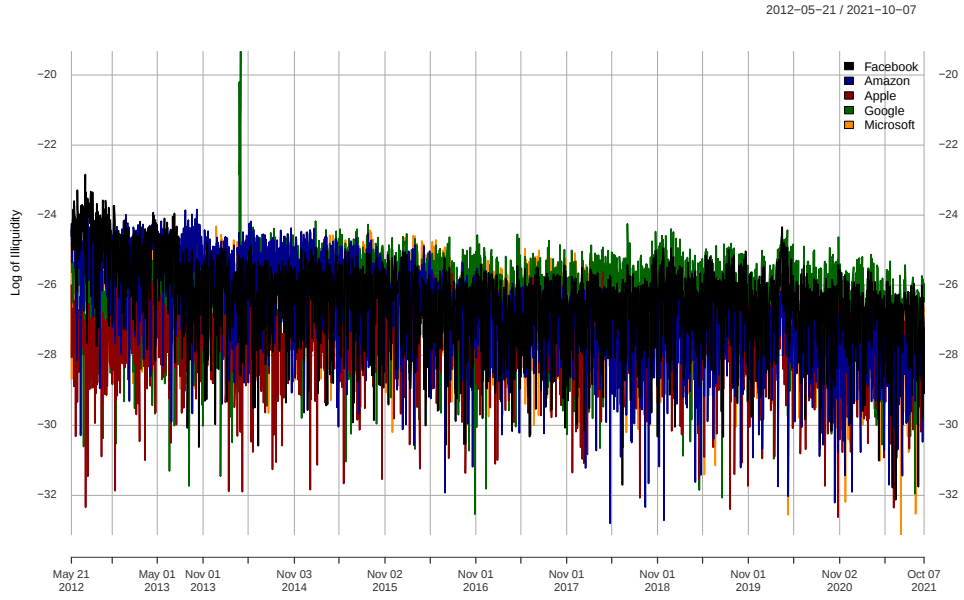
$$y_t = \alpha + \beta(t/T) + \gamma(t/T)^2 + \sum_{j=1}^5 \phi_j y_{t-j} + \varepsilon_t,$$

where the coefficients β and γ respectively capture the linear and quadratic components of the polynomial trend. The estimated coefficients with their corresponding t-statistics are provided in [Table 1](#). We observe that most of the autoregressive coefficients are statistically significant, indicating some degree of persistence in the stock illiquidity dynamics.

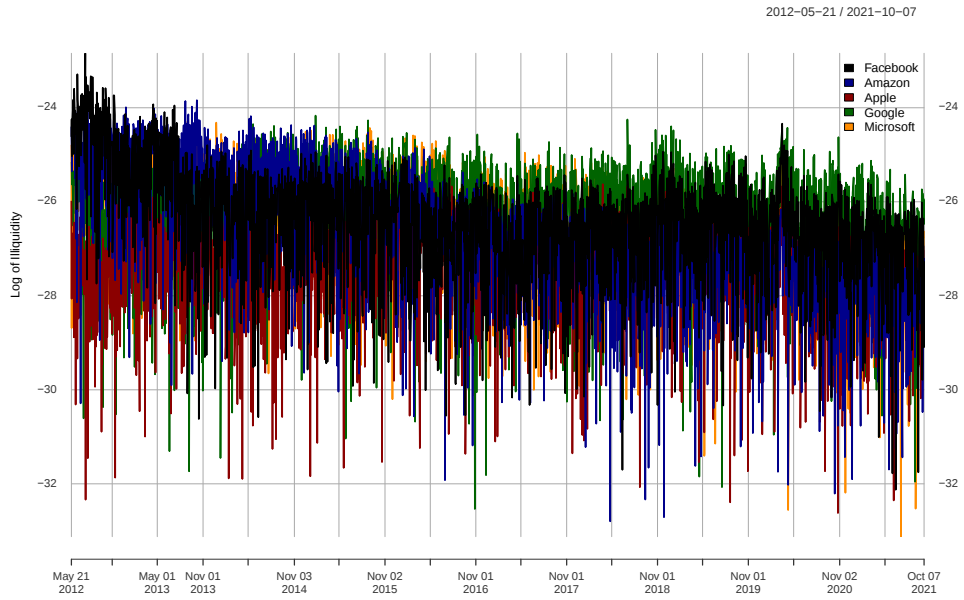
In addition, the coefficient estimates for the trend function are also significant. One exception is the quadratic term for Microsoft, meaning that this stock exhibits a linear trend over the sample period. Consistent with the visual inspection of [Figure 1](#), all the estimated polynomial trend functions are overall downward trending.

Trends in illiquidity series are not circumscribed to the Fab 5 stocks used in our illustration. To emphasize how prevalent this feature is across financial markets, [Figure 2a](#) shows the evolution of $\log \ell_t$ for the S&P500 stock market index over a longer period (1950–2021) and [Figure 2b](#) for Bitcoin-USD (2014–2021) – both exhibiting strong downward trends over their respective sample periods. Taken together, these evidence point to the existence of factors driving low-frequency variations in illiquidity dynamics in addition to higher-frequency variations.

This motivates our modelling approach for the Amihud illiquidity measure which weakens the requirement on stationarity and develop a new class of dynamic autoregressive liquidity (DARLiQ) models. A key feature of our framework is that it captures both the slow-varying long-term trend and short-run autoregressive components relevant for the modelling of illiquidity series.



(a) Original data.



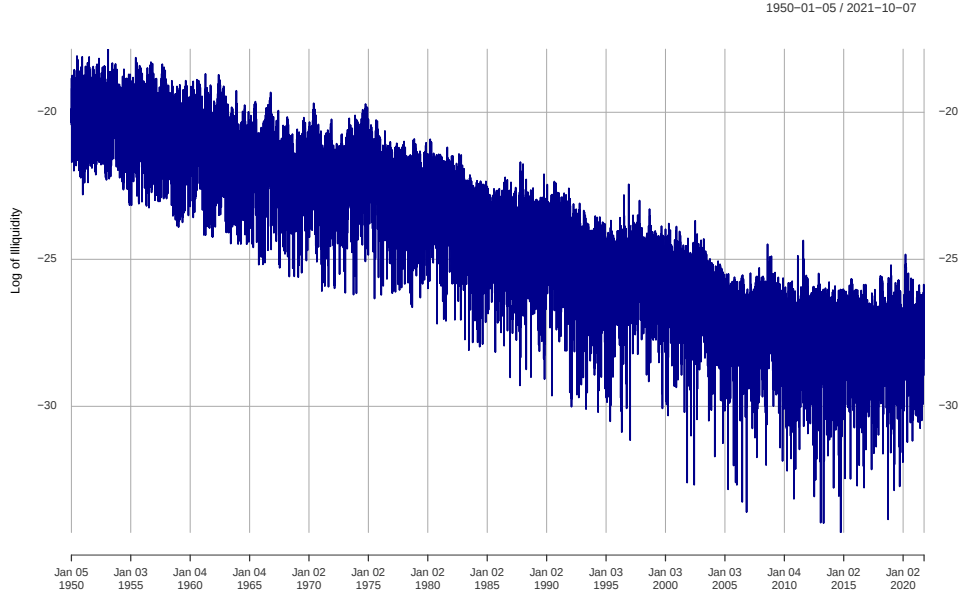
(b) Data of Google stock processed.

Figure 1: Feb 5 daily log illiquidity $-\log \ell_t$.

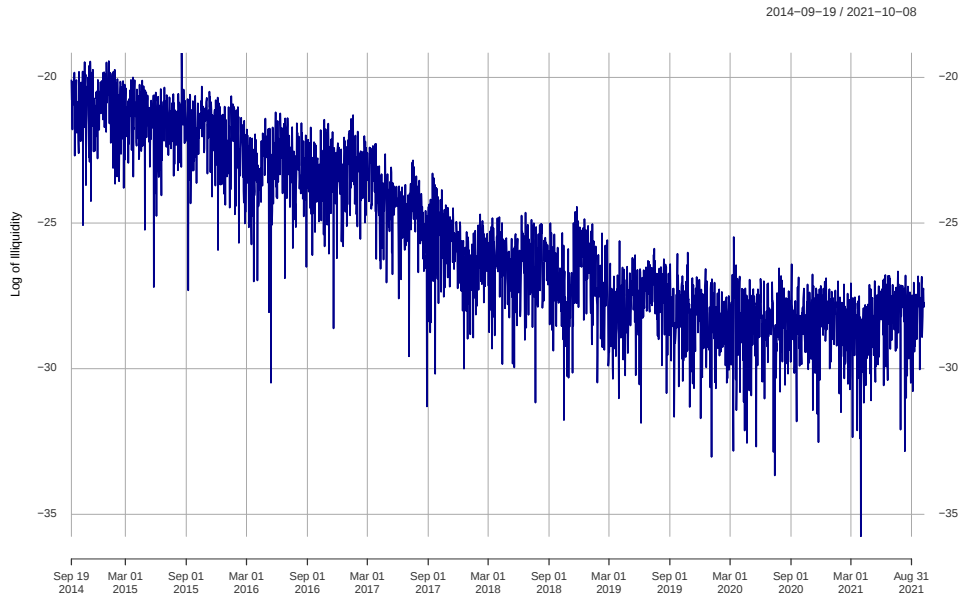
Table 1: Estimated parameters of an AR(5) with trend.

	Facebook	Amazon	Apple	Google	Microsoft
AR(1)	0.055	0.004	0.025	0.002	0.028
	(2.683)	(0.175)	(1.222)	(0.095)	(1.339)
AR(2)	0.196	0.003	0.019	0.102	0.029
	(9.720)	(0.136)	(0.929)	(4.961)	(1.398)
AR(3)	0.146	0.088	0.106	0.168	0.025
	(7.182)	(4.295)	(5.218)	(8.306)	(1.235)
AR(4)	0.170	0.097	0.027	0.113	0.075
	(8.426)	(4.730)	(1.316)	(5.506)	(3.647)
AR(5)	0.122	0.061	0.097	0.063	0.053
	(5.971)	(2.953)	(4.723)	(3.053)	(2.592)
Con	0.059	0.131	0.010	0.029	0.072
	(8.946)	(16.206)	(9.969)	(8.235)	(15.583)
t/T	-0.169	-0.297	0.027	0.049	-0.062
	(-7.263)	(-13.433)	(6.405)	(3.539)	(-5.076)
$(t/T)^2$	0.131	0.178	-0.031	-0.060	-0.001
	(6.346)	(10.402)	(-7.482)	(-4.411)	(-0.118)
Adj. R^2	0.475	0.499	0.086	0.104	0.245

Note: Models are fitted on $y_t = \ell_t \times 10^{10}$. The numbers in parentheses are the t-statistics of the corresponding parameter estimates.



(a) S&P500 index.



(b) Bitcoin asset.

Figure 2: Daily log illiquidity $-\log \ell_t$.

3 The model

We suppose that ℓ_t , $t = 1, \dots, T$ is an observed non-negative stochastic process. We assume that the observed daily series ℓ_t follows a multiplicative process as in [Engle and Gallo \(2006\)](#).

$$\ell_t = \lambda_t \zeta_t \quad (2)$$

$$\lambda_t = \omega + \sum_{j=1}^p \beta_j \lambda_{t-j} + \sum_{k=1}^q \gamma_k \ell_{t-k}, \quad (3)$$

where ζ_t is a sequence of positive random variables with conditional mean one and finite unconditional variance. Provided $\omega > 0$, the process $\lambda_t \geq \omega > 0$ with probability one (provided the initialization is also positive). Furthermore, provided $\sum_{j=1}^p \beta_j + \sum_{k=1}^q \gamma_k < 1$, the process ℓ_t is stationary in mean and follows an ARMA(p,q) process. We may further assume that ζ_t has constant conditional variance denoted σ_ζ^2 , in which case $E(\ell_t | \mathcal{F}_{t-1}) = \lambda_t$ and $\text{var}(\ell_t | \mathcal{F}_{t-1}) = \lambda_t^2 \sigma_\zeta^2$, i.e., both z_t and $z_t^2 - 1$ are martingale difference sequences, where $z_t = (\ell_t / \lambda_t - 1) / \lambda_t \sigma_\zeta$. We may further assume that ζ_t is i.i.d. with Lebesgue density function f on the positive real line.

In practice, it may be important to account for nonstationarity or trend. We allow for a nonparametric trend, so let

$$\ell_t = g(t/T) \lambda_t \zeta_t \quad (4)$$

$$\lambda_t = \omega + \sum_{j=1}^p \beta_j \lambda_{t-j} + \sum_{k=1}^q \gamma_k \ell_{t-k}^*, \quad (5)$$

where $g(\cdot)$ is a smooth but unknown function of rescaled time, and $\ell_t^* = \ell_t / g(t/T)$ is the rescaled liquidity. There is an identification issue because we can multiply and divide the two components g, λ by constants. We suppose that $E(\lambda_t) = 1$, which is achieved by setting $\omega = 1 - \sum_{j=1}^p \beta_j - \sum_{k=1}^q \gamma_k$. The series $\ell_t^* = \lambda_t \zeta_t$ possesses the same stationarity properties as ℓ_t from the model without a trend. Likewise, the error process ζ_t may possess a constant conditional variance or even be i.i.d. with some density function f . This is important for estimation but it may also be important for calculation of “Liquidity at Risk”, which would require some further assumption about the conditional quantiles of ζ_t . Note that in this model, the process ℓ_t actually depends on T and forms a triangular

array, $\ell_{t,T}$, but for notational economy we generally suppress this from the notation. We note that the process $\ell_{t,T}$ is locally stationary according to [Vogt \(2012\)](#). In the sequel, we suppose that $p = 1, q = 1$ for simplicity.

We may wish to consider the effects of interventions at some times $t_1, \dots, t_J$. We model temporary effects by dummy variables in the dynamic equation, that is, we let

$$\lambda_t = \omega + \beta\lambda_{t-1} + \sum_{j=1}^J \alpha_j D_{jt} + \gamma\ell_{t-1}^*,$$

where D_{jt} is one if an intervention occurs in period t_j . The null hypothesis of interest here is $\alpha_1 = \dots = \alpha_J = 0$ in which case the model collapses to Equation (5). We may allow the possibility of permanent effects by allowing the function g to be discontinuous at a known point $u_0 = t_0/T$, that is,

$$\lim_{u \uparrow u_0} = g_-(u_0), \quad \lim_{u \downarrow u_0} = g_+(u_0)$$

are both well defined but $g_-(u_0)$ may not be equal to $g_+(u_0)$. The size of the jump is the magnitude of the permanent effect (that is, the effect that remains permanently in the absence of further changes). The null hypothesis of interest here is that $g_-(u_0) = g_+(u_0)$ in which case the analysis of the trend is may exploit this property.

4 Estimation

Estimation is guided by assumptions made about the error ζ_t . The minimalist approach is to assume only that with probability one

$$E(\zeta_t - 1 | \mathcal{F}_{t-1}) = 0,$$

and $E(\zeta_t^2) = \sigma_\zeta^2 < \infty$. In that case, one can estimate the function $g(\cdot)$ and the identified parameters β, γ by conditional mean smoothing of ℓ_t and by the GMM approach. Provided that additional high level weak dependence conditions are satisfied, one can ensure a CLT for the resulting estimators. 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 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. with a Lebesgue density f . In this case, we may either assume that f is of unknown functional form or we may assume that f is parametrically specified, i.e., f_φ for some unknown shape parameters φ such as Exponential, Weibull, Gamma etc. The enlarged vector $(\beta, \gamma, \varphi)^\top$ can be estimated by MLE. In the semiparametric case, one needs also to estimate the error density $f(\cdot)$ along with the trend $g(\cdot)$ and the identified parameters. For forecasting future values of ℓ_t , one does not need the shock distribution, but predicting intervals and LAR (Liquidity at Risk) requires the estimation of some features of the error distribution.

Finally, we note that under the smoothness conditions on $g(\cdot)$,

$$\frac{\ell_{t,T}}{\ell_{t-1,T}} = (1 + O(T^{-1})) \frac{\lambda_t \zeta_t}{\lambda_{t-1} \zeta_{t-1}}$$

is approximately stationary and one could design a model specification test based on this “differencing” along the lines of [Yatchew \(2000\)](#).

4.1 Estimation based on conditional moment restriction

We have the unconditional moment restriction

$$E(\ell_t) = g(t/T).$$

We first use this condition to obtain an initial consistent estimators of g by kernel smoothing method, specifically

$$\widehat{g}(u) = \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \ell_t, \quad (6)$$

where K is a kernel function symmetric about zero supported on $[-1, 1]$ satisfying $\int K(u) du = 1$ and $\int K(u) u^2 du < \infty$, while h is a bandwidth sequence. Because of the equally spaced observations in time, the denominator of the Nadaraya-Watson estimator is unnecessary here (this may also be called the Priestley-Chao estimator). Another estimator of historical interest is the Gasser-Müller (1979) estimator for time trend. One could use local

linear or local polynomial estimators here to manage the boundary issue or alternatively modify the kernel in the boundary region $[-1, -1 + h] \cup [1 - h, 1]$. One can interpret the widely computed measure A_t defined in Equation (1) as a crude estimator of g at the appropriate time point. We define the detrended liquidity $\widehat{\ell}_t^* = \ell_t / \widehat{g}(t/T)$, $t = 1, \dots, T$.

Second, we use the GMM approach to estimate the parameters from the conditional moment restriction

$$E(\ell_t^* | \mathcal{F}_{t-1}) = \lambda_t,$$

where $\ell_t^* = \ell_t / g(t/T)$, $t = 1, \dots, T$. We work with $e_t(\theta) = \ell_t^* - \lambda_t(\theta)$, where $\theta = (\beta, \gamma)^\top$, which is a martingale difference sequence at the true parameter values $\beta = \beta_0, \gamma = \gamma_0$. In practice, we define $\widehat{e}_t(\theta) = \widehat{\ell}_t^* - \widehat{\lambda}_t(\theta)$ and

$$\widehat{\lambda}_t(\theta) = 1 - \beta - \gamma + \beta \lambda_{t-1} + \gamma \widehat{\ell}_{t-1}^*.$$

Then we define $\rho_t(\theta, \widehat{g}) = z_{t-1}(\widehat{\ell}_t^* - \widehat{\lambda}_t(\theta))$ and

$$M_T(\theta, \widehat{g}) = \frac{1}{T} \sum_{t=1}^T \rho_t(\theta, \widehat{g})$$

$$\widehat{\theta}_{GMM} = \arg \min_{\theta \in \Theta} \|M_T(\theta, \widehat{g})\|_W,$$

where W is a weighting matrix, while $z_t \in \mathcal{F}_t$ are instruments. This provides initial consistent estimators of θ . One can optimize the GMM procedure by choosing the instruments and weight matrix optimally, but we shall not pursue this here.

Given consistent estimates of θ , one can improve the estimate of g . Note that

$$E\left(\frac{\ell_t}{\lambda_t}\right) = g(t/T),$$

which provides an alternative local moment condition for estimation, that is, we let

$$\widetilde{g}(u) = \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{\ell_t}{\widehat{\lambda}_t}, \quad (7)$$

where $\widehat{\lambda}_t = \widehat{\lambda}_t(\widehat{\theta}_{GMM}, \widehat{g})$ are estimated in the previous procedure. We may repeat to update the estimates of θ .

The other approach to this is to use profiling, that is, for given θ , we estimate $g_\theta(u)$ by smoothing $\ell_t / \lambda_t(\theta, g_\theta)$ against time and then we optimize the profiled GMM objective function. This approach is more time consuming and we prefer the direct approach here.

4.2 Estimation based on i.i.d. assumption

In this case, we assume that the error is i.i.d. with mean one and density f . We consider several cases. First, where f is known completely. Second, where f is known up to a vector of parameters φ . Third, where f is of unknown form. [Drost and Klaassen \(1997\)](#) suppose that ζ_t is a scale random variable with $\zeta_t = \sigma_\zeta z_t + 1$, where z_t is a standardized random shock with mean zero and variance one and density f_0 . The density f_0 may be treated parametrically, i.e., depending on unknown shape parameters φ or it may be treated nonparametrically.

In the case where f is treated parametrically, one may prefer to work directly with ζ_t and parameterize f . In that case, we have a semiparametric model with parameters $\theta = (\beta, \gamma, \varphi^\top)^\top$ and unknown functions $\tau(\cdot) = g(\cdot)$. In the case where f is of unknown form, we have a semiparametric model with parameters $\theta = (\beta, \gamma)^\top$ and unknown functions $\tau(\cdot) = (f(\cdot), g(\cdot))$. In that case, we can equivalently write the model in terms of parameters $\theta = (\beta, \gamma, \sigma_\zeta^2)^\top$ and unknown functions $\tau(\cdot) = (f_0(\cdot), g(\cdot))$.

4.2.1 Parametric density case

Suppose that f depends on some unknown parameters φ , denoted as f_φ . If $g(\cdot)$ were known, the log likelihood function is apart from a term to do with $g(\cdot)$ that does not depend on parameters equal to

$$L(\theta, \varphi | \ell_1, \dots, \ell_T) = - \sum_{t=1}^T \ln \lambda_t(\theta) + \sum_{t=1}^T \ln f_\varphi(\zeta_t(\theta)),$$

$$\zeta_t(\theta) = \frac{\ell_t}{\lambda_t(\theta)g(t/T)}.$$

In practice, given a consistent estimate of $g(\cdot)$, we maximize an estimated version of this likelihood. In fact, in the semiparametric model, the efficient score functions (derived in

the appendix) for θ, φ in the presence of unknown $g(\cdot)$ are

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

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

To obtain fully efficient estimates of θ, φ , we use one-step updating from initial root-T consistent estimates. Denote $\eta = (\theta, \varphi)$ and $\tilde{\eta} = (\tilde{\theta}^\top, \tilde{\varphi}^\top)^\top$ and let $\ell_{\eta t}^* = (\ell_{\theta t}^*, \ell_{\varphi t}^*)^\top$, then let

$$\tilde{\eta} = \hat{\eta} + \left(\frac{1}{T} \sum_{t=1}^T \ell_{\eta t}^*(\hat{\theta}, \hat{\varphi}, \hat{g}) \ell_{\eta t}^*(\hat{\theta}, \hat{\varphi}, \hat{g})^\top \right)^{-1} \frac{1}{T} \sum_{t=1}^T \ell_{\eta t}^*(\hat{\theta}, \hat{\varphi}, \hat{g}) \quad (10)$$

$$\begin{aligned} \ell_{\theta t}^*(\hat{\theta}, \hat{\varphi}, \hat{g}) &= \hat{s}_2(\hat{\zeta}_t) \frac{1}{\hat{\lambda}_t} \left(\frac{\partial \hat{\lambda}_t}{\partial \theta} - \frac{\frac{1}{T} \sum_{t=1}^T \frac{\partial \log \hat{\lambda}_t}{\partial \theta} \frac{1}{\hat{\lambda}_t}}{\frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{\lambda}_t^2}} \right) \\ \ell_{\varphi t}^*(\hat{\theta}, \hat{\varphi}, \hat{g}) &= \frac{\partial f_{\hat{\varphi}}(\hat{\zeta}_t) / \partial \varphi}{f_{\hat{\varphi}}(\hat{\zeta}_t)} - \frac{\frac{1}{T} \sum_{t=1}^T \frac{\partial f_{\hat{\varphi}}(\hat{\zeta}_t) / \partial \varphi}{f_{\hat{\varphi}}(\hat{\zeta}_t)} \hat{s}_2(\hat{\zeta}_t) \frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{\lambda}_t}}{\frac{1}{T} \sum_{t=1}^T \hat{s}_2^2(\hat{\zeta}_t) \frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{\lambda}_t^2}} \hat{s}_2(\hat{\zeta}_t) \frac{1}{\hat{\lambda}_t} \\ \hat{s}_2(\zeta) &= - \left(1 + \zeta \frac{f'_{\hat{\varphi}}(\zeta)}{f_{\hat{\varphi}}(\zeta)} \right), \quad \hat{\lambda}_t = 1 - \hat{\beta} - \hat{\gamma} + \hat{\beta} \hat{\lambda}_{t-1} + \hat{\gamma} \frac{\ell_{t-1}}{\hat{g}((t-1)/T)}. \end{aligned}$$

Note that we may obtain initial consistent estimates of φ by the method of moments. For example, in the gamma case, parameterized to have mean one, the parameter φ can be estimated as one over the variance.

Under the i.i.d. structure, one can also improve the estimation of g by using local likelihood. 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

$$\sum_{t=1}^T K_h(t/T - u) \left(\ln f(\zeta_t(g)) - \ln g \right), \quad (11)$$

$$\zeta_t(g) = \frac{\ell_t}{\lambda_t g}, \quad t = 1, \dots, T, \quad (12)$$

with respect to the parameter $g \in \mathbb{R}_+$. Following [Fan and Chen \(1999\)](#), we may update the estimator of g by

$$\tilde{g}_{LL}(u) = \hat{g}(u) - \hat{L}_{gg}^{-1}(\hat{g}(u); u) \hat{L}_g(\hat{g}(u); u), \quad (13)$$

where $\hat{L}_g(g; u) = \partial \hat{L}(g; u) / \partial g$ and $\hat{L}_{gg}(g; u) = \partial^2 \hat{L}(g; u) / \partial g^2$ with

$$\hat{L}(g; u) = \sum_{t=1}^T K_h(t/T - u) \left(\ln f_{\hat{\varphi}}(\tilde{\zeta}_t(g)) - \ln g \right) \quad (14)$$

$$\tilde{\zeta}_t(g) = \frac{\ell_t}{g \lambda_t(\hat{\theta}, \hat{g})}, \quad t = 1, \dots, T. \quad (15)$$

4.2.2 Nonparametric density case

Suppose we have initial consistent estimators of $\theta, g(\cdot)$. Then, one can estimate the density function f using the residuals

$$\hat{\zeta}_t = \frac{\ell_t}{\hat{g}(t/T) \hat{\lambda}_t(\hat{\theta})}, \quad t = 1, \dots, T.$$

In particular, the kernel estimator of the density is given by

$$\hat{f}(\zeta) = \frac{1}{T} \sum_{t=1}^T K_{h_f}(\hat{\zeta}_t - \zeta),$$

where h_f is another bandwidth sequence.

Next, one may proceed in a fourth step to improve estimates of all parameters and the function g using the estimated density $\hat{f}$ as the basis of a likelihood estimation. The efficient score function (derived in the appendix) for θ in the semiparametric model with unknown f, g is

$$\begin{aligned} L_{\theta}^{**} &= \sum_{t=1}^T \ell_{\theta t}^{**}, \\ \ell_{\theta t}^{**} &= s_2(\zeta_t) \left(\left(\frac{\partial \log \lambda_t}{\partial \theta} - E \left(\frac{\partial \log \lambda_t}{\partial \theta} \right) - \left(\frac{1}{\lambda_t} - E \left(\frac{1}{\lambda_t} \right) \right) \frac{E \left[\frac{\partial \log \lambda_t}{\partial \theta} \frac{1}{\lambda_t} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) \right) \\ &\quad + \frac{\zeta_t - 1}{\sigma_{\zeta}^2} \left(E \left(\frac{\partial \log \lambda_t}{\partial \theta} \right) - \frac{E \left(\frac{1}{\lambda_t} \right) E \left[\frac{\partial \log \lambda_t}{\partial \theta} \frac{1}{\lambda_t} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right). \end{aligned}$$

Efficient estimation can be conducted by two-step estimation based on initial consistent estimates of θ, f, g :

$$\tilde{\theta} = \hat{\theta} + \left(\frac{1}{T} \sum_{t=1}^T \ell_{\theta t}^{**}(\hat{\theta}, \hat{f}, \hat{g}) \ell_{\theta t}^*(\hat{\theta}, \hat{f}, \hat{g})^\top \right)^{-1} \frac{1}{T} \sum_{t=1}^T \ell_{\theta t}^{**}(\hat{\theta}, \hat{f}, \hat{g}) \quad (16)$$

$$\begin{aligned} \ell_{\theta t}^{**}(\hat{\theta}, \hat{f}, \hat{g}) = \hat{s}_2(\hat{\zeta}_t) & \left(\left(\frac{\partial \log \hat{\lambda}_t}{\partial \theta} - \frac{1}{T} \sum_{t=1}^T \frac{\partial \log \hat{\lambda}_t}{\partial \theta} - \left(\frac{1}{\hat{\lambda}_t} - \frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{\lambda}_t} \right) \frac{\frac{1}{T} \sum_{t=1}^T \frac{\partial \log \hat{\lambda}_t}{\partial \theta} \frac{1}{\hat{\lambda}_t}}{\frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{\lambda}_t^2}} \right) \right) \\ & + \frac{\hat{\zeta}_t - 1}{\hat{\sigma}_\zeta^2} \left(\frac{1}{T} \sum_{t=1}^T \frac{\partial \log \hat{\lambda}_t}{\partial \theta} - \frac{\frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{\lambda}_t} \frac{1}{T} \sum_{t=1}^T \frac{\partial \log \hat{\lambda}_t}{\partial \theta} \frac{1}{\hat{\lambda}_t}}{\frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{\lambda}_t^2}} \right). \end{aligned}$$

where $\hat{\sigma}_\zeta^2 = \sum_{t=1}^T (\hat{\zeta}_t - \bar{\zeta})^2 / T$ with $\bar{\zeta} = \sum_{t=1}^T \hat{\zeta}_t / T$.

In practice, we may update the estimator of g by the one-step improvement

$$\tilde{g}_{LL}(u) = \hat{g}(u) - \hat{L}_{gg}^{-1}(\hat{g}(u); u) \hat{L}_g(\hat{g}(u); u),$$

where $\hat{L}_g(g; u) = \partial \hat{L}(g; u) / \partial g$ and $\hat{L}_{gg}(g; u) = \partial^2 \hat{L}(g; u) / \partial g^2$ with

$$\hat{L}(g; u) = \sum_{t=1}^T K_h(t/T - u) \left(\ln \hat{f}(\tilde{\zeta}_t(g)) - \ln g \right), \quad (17)$$

$$\tilde{\zeta}_t(g) = \frac{\ell_t}{g \hat{\lambda}_t(\hat{\theta})}, \quad t = 1, \dots, T. \quad (18)$$

These procedures can be iterated, that is, given consistent initial estimators of g, θ , we estimate f . Then we use the estimated f to update our estimate of θ taking the initial estimator of g , then we update our estimator of g using the estimated f and updated θ . We can continue this operation until it meets the defined criteria of convergence.

4.3 Sieve semiparametric estimation procedure

An alternative approach is based on the sieve method, [Chen \(2007\)](#). The advantage of this method is that it only requires a single optimization, albeit one with many parameters to choose. We just consider the cases where the error ζ_t is i.i.d. with mean one and

density f : (a) f is parametric with f_φ that reflects the unit mean constraint; (b) f is nonparametric. We suppose in both cases that

$$\ln g(u) = \sum_{j=0}^{\infty} \psi_j H_j(u)$$

for some basis functions H_j defined on $[0, 1]$. In case (b) we will also consider

$$\ln f(v) = \sum_{j=0}^{\infty} \alpha_j \Gamma_j(v)$$

for some basis functions Γ_j defined on $\mathbb{R}_+$.

We first consider case (a) where f is parametric (and φ is a scalar). In that case, the sieve log likelihood for $\eta = (\theta^\top, \varphi, \psi^\top)^\top$ is

$$L(\eta | \ell_1, \dots, \ell_T) = - \sum_{t=1}^T \ln \lambda_t(\theta) - \sum_{t=1}^T \sum_{j=0}^M \psi_j H_j(t/T) + \sum_{t=1}^T f_\varphi(\zeta_t(\theta, \psi)),$$

$$\zeta_t(\theta, \psi) = \frac{\ell_t}{\lambda_t(\theta) \exp \left(\sum_{j=0}^M \psi_j H_j(u) \right)},$$

where $M = M(T)$ is the truncation parameter. We just maximize this criterion function with respect to $\eta \in \mathbb{R}^{M+3}$. Under some restrictions, the estimates $\hat{\theta}, \hat{\varphi}$ of θ, φ are consistent and asymptotically normal with mean zero and finite variance, while we can show that the estimated function

$$\hat{g}(u) = \exp \left(\sum_{j=0}^M \hat{\psi}_j H_j(u) \right)$$

is also consistent and asymptotically normal at a slower rate. Specifically, the standard errors are calculated as follows. Define the matrices

$$\hat{\mathcal{I}}_{\eta\eta} = \frac{\partial^2 L(\hat{\eta} | \ell_1, \dots, \ell_T)}{\partial \eta \partial \eta^\top}, \quad \hat{\mathcal{I}}_{\eta\eta}^{-1} = \begin{pmatrix} \hat{\mathcal{I}}^{\theta\theta} & \hat{\mathcal{I}}^{\theta\varphi} & \hat{\mathcal{I}}^{\theta\psi} \\ \hat{\mathcal{I}}^{\varphi\theta} & \hat{\mathcal{I}}^{\varphi\varphi} & \hat{\mathcal{I}}^{\varphi\psi} \\ \hat{\mathcal{I}}^{\psi\theta} & \hat{\mathcal{I}}^{\psi\varphi} & \hat{\mathcal{I}}^{\psi\psi} \end{pmatrix}.$$

Then it can be shown that

$$\left(\hat{\mathcal{I}}^{\theta\theta} \right)^{-1/2} \left(\hat{\theta} - \theta \right) \Longrightarrow N(0, I_2), \quad \left(\hat{\mathcal{I}}^{\varphi\varphi} \right)^{-1/2} \left(\hat{\varphi} - \varphi \right) \Longrightarrow N(0, 1), \quad (19)$$

$$\left(H^\top \widehat{\mathcal{I}}^{\psi\psi} H\right)^{-1/2} \exp(-\widehat{g}(u)) (\widehat{g}(u) - g(u)) \implies N(0, 1), \quad (20)$$

where $H = (H_1, \dots, H_M)^\top$ and we have suppressed the dependence on u .

We now turn to case (b) where f is treated nonparametrically. In this case we also need to impose the restrictions that f is a density and that ζ has mean one. In this case, the restricted sieve log likelihood for $\eta = (\theta^\top, \psi^\top, \alpha^\top, \delta)^\top$ is

$$\begin{aligned} L(\theta, \psi, \alpha, \delta | \ell_1, \dots, \ell_T) = & - \sum_{t=1}^T \ln \lambda_t(\theta) - \sum_{t=1}^T \sum_{j=0}^M \psi_j H_j(t/T) + \sum_{t=1}^T \sum_{j=0}^N \alpha_j \Gamma_j(\zeta_t(\theta, \psi)) \\ & - \ln \left(\int \exp \left(\sum_{j=0}^N \alpha_j \Gamma_j(z) \right) dz \right) - \delta \left(\int z \exp \left(\sum_{j=0}^N \alpha_j \Gamma_j(z) \right) dz - 1 \right), \end{aligned}$$

where $\zeta_t(\theta, \psi)$ is as defined above and δ is a Lagrange multiplier parameter that imposes the unit mean restriction on f . We find the solution to the first order condition with respect to $\eta \in \mathbb{R}^{M+N+2}$. Define for this η the matrices

$$\widehat{\mathcal{I}}_{\eta\eta} = \frac{\partial^2 L(\widehat{\eta} | \ell_1, \dots, \ell_T)}{\partial \eta \partial \eta^\top}, \quad \widehat{\mathcal{I}}_{\eta\eta}^{-1} = \begin{pmatrix} \widehat{\mathcal{I}}^{\theta\theta} & \widehat{\mathcal{I}}^{\theta\psi} & \widehat{\mathcal{I}}^{\theta\alpha} & \widehat{\mathcal{I}}^{\theta\delta} \\ & \widehat{\mathcal{I}}^{\psi\psi} & \widehat{\mathcal{I}}^{\psi\alpha} & \widehat{\mathcal{I}}^{\psi\delta} \\ & & \widehat{\mathcal{I}}^{\alpha\alpha} & \widehat{\mathcal{I}}^{\alpha\delta} \\ & & & \widehat{\mathcal{I}}^{\delta\delta} \end{pmatrix}.$$

Then it can be shown that

$$\left(\widehat{\mathcal{I}}^{\theta\theta}\right)^{-1/2} \left(\widehat{\theta} - \theta\right) \implies N(0, I_2), \quad (21)$$

$$\left(H^\top \widehat{\mathcal{I}}^{\psi\psi} H\right)^{-1/2} \exp(-\widehat{g}(u)) (\widehat{g}(u) - g(u)) \implies N(0, 1), \quad (22)$$

$$\left(\Gamma^\top \widehat{\mathcal{I}}^{\alpha\alpha} \Gamma\right)^{-1/2} \exp(-\widehat{f}(\zeta)) (\widehat{f}(\zeta) - f(\zeta)) \implies N(0, 1), \quad (23)$$

where $H = (H_1, \dots, H_M)^\top$, $\Gamma = (\Gamma_1, \dots, \Gamma_N)^\top$, and we have suppressed the dependence on u, ζ .

The difficulty with this approach is the potential high dimensional parameter space that has to be navigated when the truncation parameters M, N are large.

5 Large sample properties

The distribution theory for this requires some extension of [Hafner and Linton \(2010\)](#). The presence of two nonparametric functions g, f , albeit scalar ones, makes this a little challenging. We make some assumptions.

Definition. A triangular array process $\{X_{t,T}, t = 0, 1, 2, \dots, T = 1, 2, \dots\}$ is said to be alpha mixing if

$$\alpha(k) = \sup_{T \geq 1} \sup_{A \in \mathcal{F}_{-\infty}^n, B \in \mathcal{F}_{n+k}^\infty} |P(AB) - P(A)P(B)| \rightarrow 0, \quad (24)$$

as $k \rightarrow \infty$, where $\mathcal{F}_{-\infty}^{n,T}$ and $\mathcal{F}_{n+k,T}^\infty$ are two σ -fields generated by $\{X_{t,T}, t \leq n\}$ and $\{X_{t,T}, t \geq n+k\}$ respectively. We call $\alpha(\cdot)$ the mixing coefficient.

We suppose that ℓ_t^* is stationary and alpha mixing. This can be shown to hold under the parameter restrictions provided ζ_t . It may also hold when ζ_t itself is only described as a stationary mixing process although this can be difficult to establish. Instead, one can work with the more general near epoch dependence condition, see [Lu and Linton \(2007\)](#).

We define the long run variance for a stationary mixing process x_t as

$$\text{lrvar}(x_t) = \sum_{j=-\infty}^{\infty} \text{cov}(x_t, x_{t-j}).$$

5.1 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. We do not address the efficient use of this information but it follows from standard arguments.

5.1.1 Nonparametric trend

We first consider the estimator $\hat{g}(u)$, $u \in (0, 1)$, that is based on smoothing of the raw liquidity. Let $v_t = \lambda_t \zeta_t - 1$.

ASSUMPTION A1. Suppose that g is twice continuously differentiable at $u \in (0, 1)$.

ASSUMPTION A2. Suppose that $\{v_t\}$ is an alpha mixing sequence with $E(v_t) = 0$ and $E(|v_t|^{2+\delta}) \leq C < \infty$ for some $\delta > 0$ for $t = 1, 2, \dots$ such that

$$\sum_{k=1}^{\infty} \alpha(k)^{\frac{\delta}{2+\delta}} < \infty.$$

ASSUMPTION A3. Suppose that K is symmetric about zero with compact support $[-1, 1]$ and K is differentiable on its support.

Theorem 1. Suppose that assumptions A1-A3 hold and that $h \rightarrow 0$ and $Th \rightarrow \infty$. Then

$$\sqrt{Th} (\hat{g}(u) - g(u) - h^2 b(u)) \implies N(0, V(u)),$$

$$V(u) = g^2(u) \|K\|^2 \times \text{lrvar}(v_t).$$

This estimator is consistent and asymptotically normal with a potentially optimal rate of $T^{-2/5}$ based on the smoothness assumption. The bandwidth that achieves this rate is of order $h = T^{-1/5}$ and balances squared bias with variance. We know that $\ell_t^* = \ell_t/g(t/T) = \lambda_t \zeta_t$ is an ARMA(1,1) process with $A(L)\ell_t^* = B(L)e_t$ for some MDS shock e_t and lag polynomials A, B . In this case, the long run variance of ℓ_t^* is $\sigma_e^2(B(1)/A(1))^2$ and it could be estimated by the plug in of estimated θ from the second step.

Instead, it may be preferable to work with the refined estimator $\tilde{g}(u)$ that is based on the estimator of θ . We have for this estimator the following CLT.

Theorem 2. Suppose that assumptions A1-A3 hold and that $\hat{\theta}$ is $\sqrt{T}$ -consistent. Suppose that $h \rightarrow 0$ and $Th \rightarrow \infty$. Then

$$\sqrt{Th} (\tilde{g}(u) - g(u) - h^2 b(u)) \implies N(0, V(u)),$$

$$V(u) = g^2(u) \|K\|^2 \times \text{var}(\zeta_t).$$

In this case, the limiting variance is proportional to the variance of ζ_t , which is generally smaller and easier to estimate. When ζ_t is i.i.d.,

$$E(\lambda_t^2(\zeta_t - 1)^2) = E(\lambda_t^2) \times \text{var}(\zeta_t) \geq \text{var}(\zeta_t),$$

because by assumption $E(\lambda_t) = 1$. For this estimator, consistent standard errors can be based on

$$\widehat{V}(u) = \widetilde{g}^2(u) \|K\|^2 \times \widehat{\sigma}_\zeta^2, \quad (25)$$

where $\widehat{\sigma}_\zeta^2$ is an estimator of the unconditional variance of ζ_t such as defined above.

5.1.2 Parametric Components

Let $\theta = (\beta, \gamma)^\top$ and let W be some given positive definite weighting matrix. Let

$$w_t = \lambda_t(\zeta_t - 1)z_{t-1} + \frac{1 - \beta - \gamma}{1 - \beta}(\lambda_t\zeta_t - 1)E(z_{t-1}).$$

We consider for any function g

$$M_T(\theta, g) = \frac{1}{T} \sum_{t=1}^T \rho_t(\theta, g), \quad \rho_t(\theta, g) = z_{t-1} \left(\frac{\ell_t}{g(t/T)} - \lambda_t(\theta, g) \right)$$

$$\lambda_t(\theta, g) = 1 - \beta - \gamma + \beta\lambda_{t-1} + \gamma \frac{\ell_{t-1}}{g((t-1)/T)} = \frac{1 - \beta - \gamma}{1 - \beta} + \gamma \sum_{j=1}^{\infty} \beta^{j-1} \frac{\ell_{t-j}}{g((t-j)/T)}.$$

ASSUMPTION A4. *Define:*

$$M(\theta, g) = \lim_{T \rightarrow \infty} E(M_T(\theta, g))$$

$$\Gamma_1(\theta, g_0) = \frac{\partial M(\theta, g_0)}{\partial \theta}$$

$$\Gamma_2(\theta, g_0) \circ (g - g_0) = \frac{\partial}{\partial \tau} M(\theta, g_0 + \tau(g - g_0)),$$

which are assumed to exist in all directions θ, g .

ASSUMPTION A5. *For all $\delta > 0$, there is a $\epsilon > 0$ such that*

$$\inf_{\|\theta - \theta_0\| > \delta} \|M(\theta, g_0)\| \geq \epsilon.$$

Uniformly for all $\theta \in \Theta$, the function $M(\theta, g)$ is continuous in g (with the respect to the L_2 metric) at $g = g_0$. Furthermore,

$$\sup_{\theta \in \Theta, \|g - g_0\| \leq \delta_T} \|M_T(\theta, g) - M(\theta, g_0)\| = o_P(1).$$

ASSUMPTION A6. For all sequences of positive numbers $\delta_T \rightarrow 0$

$$\begin{aligned} \sup_{\|\theta - \theta_0\| \leq \delta_T, \|g - g_0\| \leq \delta_T} \delta_T^{-2} \|M(\theta, g) - M(\theta, g) - \Gamma_2(\theta, g_0) \circ (g - g_0)\| &\leq C \\ \sup_{\|\theta - \theta_0\| \leq \delta_T, \|g - g_0\| \leq \delta_T} \delta_T^{-1} \|\Gamma_2(\theta, g_0) \circ (g - g_0) - \Gamma_2(\theta_0, g_0) \circ (g - g_0)\| &= o(1) \\ \sup_{\|\theta - \theta_0\| \leq \delta_T, \|g - g_0\| \leq \delta_T} \sqrt{T} \|M_T(\theta, g) - M(\theta, g) - M_T(\theta_0, g_0)\| &= o_P(1) \end{aligned}$$

Theorem 3. Suppose that Assumptions A1-A6 hold. Then as $T \rightarrow \infty$

$$\begin{aligned} \sqrt{T} (\hat{\theta} - \theta) &\implies N(0, V) \\ \Omega &= \lim_{T \rightarrow \infty} \text{var} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T w_t \right), \quad \Gamma = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left(\frac{\partial e_t(\theta_0)}{\partial \theta} z_{t-1} \right) \\ V &= (\Gamma^\top W \Gamma)^{-1} (\Gamma^\top W \Omega W \Gamma) (\Gamma^\top W \Gamma)^{-1}. \end{aligned}$$

In general, the asymptotic variance will depend on the long run variance of the process w_t and so inference procedures are complicated by that.

5.2 I.i.d. Shocks

We suppose here that ζ_t is i.i.d. with mean one and density f . Define the so-called Fisher score functions and informations

$$s_1(\zeta) = \frac{f'(\zeta)}{f(\zeta)}, \quad s_2(\zeta) = 1 + \zeta \frac{f'(\zeta)}{f(\zeta)} \quad (26)$$

$$I_j(f) = \int s_j^2(\zeta) f(\zeta) d\zeta, \quad j = 1, 2. \quad (27)$$

We suppose that we have initial consistent estimators of $g(\cdot), \theta$ available from the GMM procedure described above, say.

5.2.1 Parametric Density

In the case where f is parametrically specified with parameters φ , the model is semiparametric with parameters $\eta = (\theta^\top, \varphi)^\top$ and unknown function g . We first consider the local likelihood estimator of the trend function. We note that regarding the estimation of $g(\cdot)$ it may be assumed that the error density is known, whether this is estimated parametrically or nonparametrically, [Linton and Xiao \(2001\)](#).

Theorem 4. *Suppose that assumptions A1-A3 hold and that $\hat{\theta}$ is $\sqrt{T}$ -consistent. Suppose that $h \rightarrow 0$ and $Th \rightarrow \infty$. The local likelihood estimator has for some bias $b(u)$,*

$$\sqrt{Th} (\tilde{g}_{LL}(u) - g(u) - h^2 b(u)) \implies N(0, V(u))$$

$$V(u) = \|K\|^2 I_2^{-1}(f) g(u)^2.$$

To construct pointwise confidence bands we use:

$$\begin{aligned} \tilde{V}(u) &= \|K\|^2 I_2^{-1}(f_{\tilde{\varphi}}) \tilde{g}(u)^2, \\ I_2(f_{\tilde{\varphi}}) &= \frac{1}{T} \sum_{t=1}^T \left(1 + \zeta \frac{f'_{\tilde{\varphi}}(\tilde{\zeta}_t)}{f_{\tilde{\varphi}}(\tilde{\zeta}_t)} \right)^2, \end{aligned}$$

where $\tilde{\zeta}_t$ are the estimated residuals.

We next turn to the properties of the estimated parametric components. Define the efficient information matrix:

$$\mathcal{I}_{\eta\eta}^* = \begin{pmatrix} \mathcal{I}_{\theta\theta}^* & \mathcal{I}_{\theta\varphi}^* \\ \mathcal{I}_{\varphi\theta}^* & \mathcal{I}_{\varphi\varphi}^* \end{pmatrix}, \quad \mathcal{I}_{\theta\theta}^* = E(\ell_{\theta t}^* \ell_{\theta t}^{*\top}), \quad \mathcal{I}_{\varphi\varphi}^* = E(\ell_{\varphi t}^* \ell_{\varphi t}^{*\top}), \quad \mathcal{I}_{\theta\varphi}^* = E(\ell_{\theta t}^* \ell_{\varphi t}^{*\top}).$$

Theorem 5. *Suppose that Assumptions A1-A6 hold. Then as $T \rightarrow \infty$*

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

Furthermore, the asymptotic variance may be estimated by $\tilde{\mathcal{I}}^{*-1}$, where

$$\tilde{\mathcal{I}}_{\eta\eta}^* = \frac{1}{T} \sum_{t=1}^T \ell_{\eta t}^*(\tilde{\theta}, \tilde{\varphi}, \tilde{g}) \ell_{\eta t}^{*\top}(\tilde{\theta}, \tilde{\varphi}, \tilde{g}).$$

5.2.2 Nonparametric Density Case

In the case where f is nonparametrically specified, the model is semiparametric with parameters θ and unknown functions g, f . First, the large sample properties of the estimated g are as if f is known. We turn to the properties of the estimated parametric components.

Define the efficient information matrix

$$\mathcal{I}_{\theta\theta}^{**} = E \left(\ell_{\theta t}^{**} \ell_{\theta t}^{**\top} \right).$$

Theorem 6. *Suppose that Assumptions A1-A6 hold. Then as $T \rightarrow \infty$*

$$\sqrt{T} \left(\tilde{\theta} - \theta \right) \Rightarrow N(0, \mathcal{I}_{\theta\theta}^{*-1}).$$

Furthermore, the asymptotic variance may be estimated by $\tilde{\mathcal{I}}^{*-1}$, where

$$\tilde{\mathcal{I}}_{\theta\theta}^{**} = \frac{1}{T} \sum_{t=1}^T \ell_{\theta t}^{**}(\tilde{\theta}, \tilde{f}, \tilde{g}) \ell_{\theta t}^{**\top}(\tilde{\theta}, \tilde{f}, \tilde{g}).$$

The presence of unknown τ does not affect the efficiency of θ once the scale σ_ζ^2 is accounted for, but the scale does affect the achievable efficiency.

6 Testing for Temporary and Permanent Shifts

We can estimate the function g allowing for a discontinuity at the point $u_0 = t_0/T$ by considering

$$\hat{g}^+(u) = \frac{1}{T} \sum_{t=1}^T K_h^+(t/T - u) \ell_t, \quad \hat{g}^-(u) = \frac{1}{T} \sum_{t=1}^T K_h^-(t/T - u) \ell_t,$$

where K^+ is a kernel supported on $[0, 1]$ with $\int K^+(u) du = 1$ and $\int K^+(u) u du = 0$ and K^- is a kernel supported on $[-1, 0]$ with $\int K^-(u) du = 1$ and $\int K^-(u) u du = 0$. We may test for the presence of a discontinuity by computing

$$\tau(u_0) = \sqrt{Th} \frac{\hat{g}^+(u_0) - \hat{g}^-(u_0)}{\sqrt{\hat{\sigma}^{2+}(u_0) \|K^+\|^2 + \hat{\sigma}^{2-}(u_0) \|K^-\|^2}},$$

where $\widehat{\sigma}^{2\pm}(u_0) ||K^\pm||^2 / Th$ are estimates of the asymptotic variance of $\widehat{g}^\pm(u_0)$. In general,

$$\widehat{\sigma}^{2\pm}(u_0) = \widehat{g}^\pm(u_0)^2 \times \widehat{\text{Irvar}}(\lambda_t \zeta_t),$$

because $\ell_t - g(t/T) = g(t/T)(\lambda_t \zeta_t - 1)$ has smoothly varying variance that is exactly $g^2(t/T)$ and because the series $\lambda_t \zeta_t$ is in general weakly dependent. On the other hand if we work with the improved estimator that works with a smooth of $\ell_t / \widehat{\lambda}_t$, we can assume that the error process is a MDS and so the variance of the estimator is proportional to $g^\pm(t/T)^2 \sigma_\zeta^2$. That is, we may define

$$\widetilde{g}^+(u) = \frac{1}{T} \sum_{t=1}^T K_h^+(t/T - u) \frac{\ell_t}{\widehat{\lambda}_t}, \quad \widetilde{g}^-(u) = \frac{1}{T} \sum_{t=1}^T K_h^-(t/T - u) \frac{\ell_t}{\widehat{\lambda}_t}.$$

In this case, we may choose

$$\widetilde{\sigma}^{2\pm}(u_0) = \widetilde{g}^\pm(u_0)^2 \times \widehat{\sigma}_\zeta^2.$$

Likewise for the local likelihood estimator, but this also takes care of the error shape and heteroskedasticity. Given the studentized statistic, $\tau(u_0)$ we compare this with the standard normal distribution as in [Delgado and Hidalgo \(2000\)](#). Under the null hypothesis this should lie between $\pm z_{\alpha/2}$ with probability $1 - \alpha$.

We also consider how to include a control group to eliminate common trends at the change time. This amounts to a diff in diff test. Specifically, suppose that we have a “treatment” stock labelled with an S subscript and a “control” stock labelled with an C subscript. We suppose that model (4) holds for both stocks and that ζ_{St} and ζ_{Ct} may be correlated. We define the diff-in-diff statistic as

$$\tau_{did}(u_0) = \sqrt{Th} \frac{(\widetilde{g}_S^+(u_0) - \widetilde{g}_S^-(u_0)) - (\widetilde{g}_C^+(u_0) - \widetilde{g}_C^-(u_0))}{\sqrt{\left(\widetilde{\sigma}_S^{2+}(u_0) + \widetilde{\sigma}_C^{2+}(u_0) - 2\widetilde{\sigma}_{S,C}^+(u_0)\right) ||K^+||^2 + \left(\widetilde{\sigma}_S^{2-}(u_0) + \widetilde{\sigma}_C^{2-}(u_0) - 2\widetilde{\sigma}_{S,C}^-(u_0)\right) ||K^-||^2}},$$

where:

$$\begin{aligned} \widetilde{\sigma}_{S,C}^\pm(u) &= \widetilde{g}_S^\pm(u_0) \widetilde{g}_C^\pm(u_0) \times \widehat{\sigma}_{\zeta_S, \zeta_C}, \\ \widehat{\sigma}_{\zeta_S, \zeta_C} &= \frac{1}{T} \sum_{t=1}^T \left(\widehat{\zeta}_{St} - \bar{\zeta}_S \right) \left(\widehat{\zeta}_{Ct} - \bar{\zeta}_C \right). \end{aligned}$$

What happens if we reject this test at some point $u_0 = t_0/T$? Note that the above procedures (estimation of θ) can allow $\widehat{g}(u)$ to be inconsistent at a finite number of points, because the estimators of θ, f are averages of $\widehat{g}(t/T)$ over all points $t = 1, \dots, T$. However, the rate of convergence may be affected, since the jump points will contribute a bias of order $O(h)$ rather than the usual order $O(h^2)$. Therefore, one should choose a smaller bandwidth such that $\sqrt{T}h \rightarrow 0$ instead of $\sqrt{T}h^2 \rightarrow 0$. Having identified the jump points one can use this information in the rescaling. We suppose in general that the function g is continuous from the right with limits from the left. In that case for $t \geq t_0$ we may normalize by $\widehat{g}^+(u)$ and for $t < t_0$ we normalize by $\widehat{g}^-(u)$. That is, we let

$$\widehat{\ell}_t^* = \begin{cases} \frac{\ell_t}{\widehat{g}^+(t/T)} & t = t_0, \dots, T \\ \frac{\ell_t}{\widehat{g}^-(t/T)} & t = 1, \dots, t_0 - 1 \end{cases},$$

and proceed to estimate the dynamic parameters as before.

We next discuss the estimation of temporary effects in the dynamic equation

$$\lambda_t = 1 - \beta - \gamma + \beta\lambda_{t-1} + \sum_{j=0}^{J-1} \alpha_j D_{jt} + \gamma \ell_{t-1}^*,$$

where J is fixed. Here, D_{jt} are dummy variables indicating times $t_0, \dots, t_{J-1}$ and we focus on the case where $t_j = t_0 + j$. In this case it is not possible to consistently estimate the parameters α_j , however, it is possible to provide a consistent test of the null hypothesis that $\alpha_1 = \dots = \alpha_J = 0$ against the general alternative even in the full semiparametric model. The efficient score function with respect to α in the semiparametric model with known f is

$$\sum_{t=1}^T \ell_{\theta t}^* = \sum_{t=1}^T s_2(\zeta_t) \frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \alpha} - \frac{\frac{1}{T} \sum_{t=1}^T E \left[\frac{\partial \lambda_t}{\partial \alpha} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right)$$

where

$$\frac{\partial \lambda_t(\theta, \alpha)}{\partial \alpha_j} = \beta \frac{\partial \lambda_{t-1}(\theta, \alpha)}{\partial \alpha_j} + D_{jt} = \begin{cases} \beta^{t-t_j} & \text{if } t \geq t_j \\ 0 & \text{if } t < t_j. \end{cases}$$

$$\frac{1}{T} \sum_{t=1}^T E \left[\frac{\partial \lambda_t}{\partial \alpha} \frac{1}{\lambda_t^2} \right] = \frac{1}{T} \sum_{t=t_j}^T \beta^{t-t_j} E \left[\frac{1}{\lambda_t^2} \right].$$

It follows that the efficient score function at $\alpha = 0$ is

$$\frac{\partial L}{\partial \alpha_j}(\theta, \sigma_\zeta^2, 0) = \sum_{t=t_j}^T s_2(\zeta_t) \frac{1}{\lambda_t(\theta, 0)} \left(\left[\beta^{t-t_j} - \frac{1}{T} \sum_{t=t_j}^T \beta^{t-t_j} \right] \right) \simeq \sum_{t=t_j}^T s_2(\zeta_t) \frac{\beta^{t-t_j}}{\lambda_t(\theta, 0)}.$$

In practice we must replace the unknown quantities by estimates. Define the test statistic (we call CAR to recognize the event study literature where this quantity originates):

$$\widehat{CAR}(\tau) = \sum_{j=0}^{\tau} \sum_{t=t_j}^T s_2(\hat{\zeta}_t) \frac{\hat{\beta}^{t-t_j}}{\lambda_t(\hat{\theta}, 0)}, \quad \tau = 0, \dots, J-1. \quad (28)$$

The test statistic does not satisfy a central limit theorem (even when θ is known) because of the summability of $\sum_{t=t_j}^T \beta^{2(t-t_j)}$ (that is, essentially only a finite number of periods matter). Nevertheless, if the distribution of ζ_t were known along with the parameter values θ , we can calculate the distribution numerically using data $w_r(\tau) = \sum_{j=1}^{\tau} \sum_{t=r}^T s_2(\zeta_t) \beta^{t-r} / \lambda_t(\theta, 0)$ for r some time before t_0 . Let F_w denote the distribution of the series $\{w_r\}$. We assume that t_0 is large, i.e., $t_0 \rightarrow \infty$ so that there is a long sample of data available before the intervention.

We compare $\widehat{CAR}(\tau)$ with the critical values $\hat{F}_{\hat{w}}^{-1}(\alpha/2)$, $\hat{F}_{\hat{w}}^{-1}(1 - \alpha/2)$, where $\hat{F}_{\hat{w}}(\cdot)$ is estimated using the data

$$\hat{w}_r(\tau) = \sum_{j=1}^{\tau} \sum_{t=r}^T s_2(\hat{\zeta}_t) \frac{\hat{\beta}^{t-r}}{\lambda_t(\hat{\theta}, 0)}, \quad r = 1, \dots, t_0 - J. \quad (29)$$

7 Risk premium

[Amihud \(2002\)](#) considers an autoregressive model for annual and monthly liquidity and then relates this to the stock risk premium. Specifically, he writes

$$E(R_{mt} - R_{ft} | liq_t^e) = a + b \times liq_t^e$$

$$liq_t = c_0 + c_1 \times liq_{t-1} + \eta_t,$$

where $liq_t^e = E(liq_t | \mathcal{F}_{t-1}) = c_0 + c_1 liq_{t-1}$ and liq_t is the annual or monthly average that we have called A_t . He also considers a specification with unexpected liquidity as a regressor where $liq_t^u = liq_t - liq_t^e$.

It seems a bit strange to wait for a whole year to update ones estimate of liquidity. We may consider the following specification for daily stock returns

$$E(R_{mt} - R_{ft} | \mathcal{F}_{t-1}) = a + b \times g(t/T) + c \times \lambda_t + d \times \zeta_t, \quad (30)$$

where λ_t is defined above. This allows the risk premium to depend on long run trend liquidity on short run predictable dynamic variation and also on unanticipated liquidity shocks.

We also consider the alternative regression with detrended equity premium, that is,

$$E(R_{mt} - R_{ft} - m(t/T) | \mathcal{F}_{t-1}) = \alpha + \gamma \times \lambda_t + \delta \times \zeta_t, \quad (31)$$

where $m(t/T) = E(R_{mt} - R_{ft})$ is the time varying unconditional equity premium. In practice we can estimate $m(\cdot)$ by kernel smoothing methods.

8 Empirical application

The ability to accurately model illiquidity series, and the availability of a framework to conduct inference on potential structural changes in their dynamics, are useful tools to investigate liquidity conditions in financial markets and their evolution over time. In our empirical application, we consider the Fab 5 tech stocks and the Bitcoin asset introduced in Section 2 to analyze their illiquidity series using our DArLiQ model. The rest of this section is organized as follows: Section 8.1 presents the data description and the key descriptive statistics of the illiquidity series. Section 8.2.1 covers the estimation results adopting a GMM approach based on moment restrictions. Sections 8.2.2 and 8.2.3 present the analysis based on a semiparametric ML estimation procedure with, respectively, parametrically specified and nonparametrically estimated error densities. Section 8.3 focuses on the detection of permanent and temporary breaks in the illiquidity series arising from stock splits. Finally, We investigate in Section 8.4 how market liquidity dynamics – particularly short-run autoregressive dynamics and unexpected liquidity shocks – impact equity risk premia.

8.1 Data description

We use historical daily return and volume data from a panel composed of the five tech stocks and the bitcoin asset to compute the Amihud illiquidity series.² The sample period starts from the date of the first available data point for each asset until October 7th, 2021. The descriptive statistics of the illiquidity series are summarized in Table 2.³ It can be observed that the Bitcoin asset is less liquid compared to the technology company stocks during this period. In addition, the illiquidity series of Bitcoin is more volatile, exhibits higher skewness and has thicker tails. We further note that the five tech companies have comparable levels of liquidity – although Apple stock is slightly more liquid than others. Moreover, the illiquidity of Facebook stock has higher skewness and thicker tails compared to the other four tech companies.

Table 2: 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

We plot in Figure 3 and Figure 4 respectively the illiquidity and log illiquidity series over the corresponding sample period for each of the six assets. To manage boundary issues, we obtain an initial consistent estimator of the trend function $g(t/T)$ using a local linear estimator. The red curves in the two figures represent respectively the estimated trend functions and their logarithms. From Figure 3, we observe that the estimated trend function $g(t/T)$ serves as a good approximation for the time-varying mean of the

²Data were retrieved from Yahoo Finance.

³To make it comparable, we use the daily Amihud illiquidity ratios in the common sample period of September 19, 2014 to October 7th, 2021 to compute the descriptive statistics for all assets.

illiquidity series.⁴ Furthermore, a strong downward trend is observed in the evolution of most illiquidity series, indicating an overall improvement in liquidity conditions over time. Lastly, it is worth noticing that a temporary worsening in liquidity conditions is occurring during significant market events such as the burst of the dot-com bubble and 2007-2009 Global Financial Crisis.

⁴Note that the trend function $g(t/T)$ is the mean level of the illiquidity ℓ_t , i.e. $E(\ell_t) = g(t/T)$, which is estimated with a local linear estimator. Therefore, $g(t/T)$ is roughly moving around the mid-level of ℓ_t but this is not the case for the log illiquidity series as $\log g(t/T)$ is higher than the mean level of $\log \ell_t$ due to a Jensen's inequality effect. In the interest of space, we will focus on the plot of the log illiquidity series hereafter.

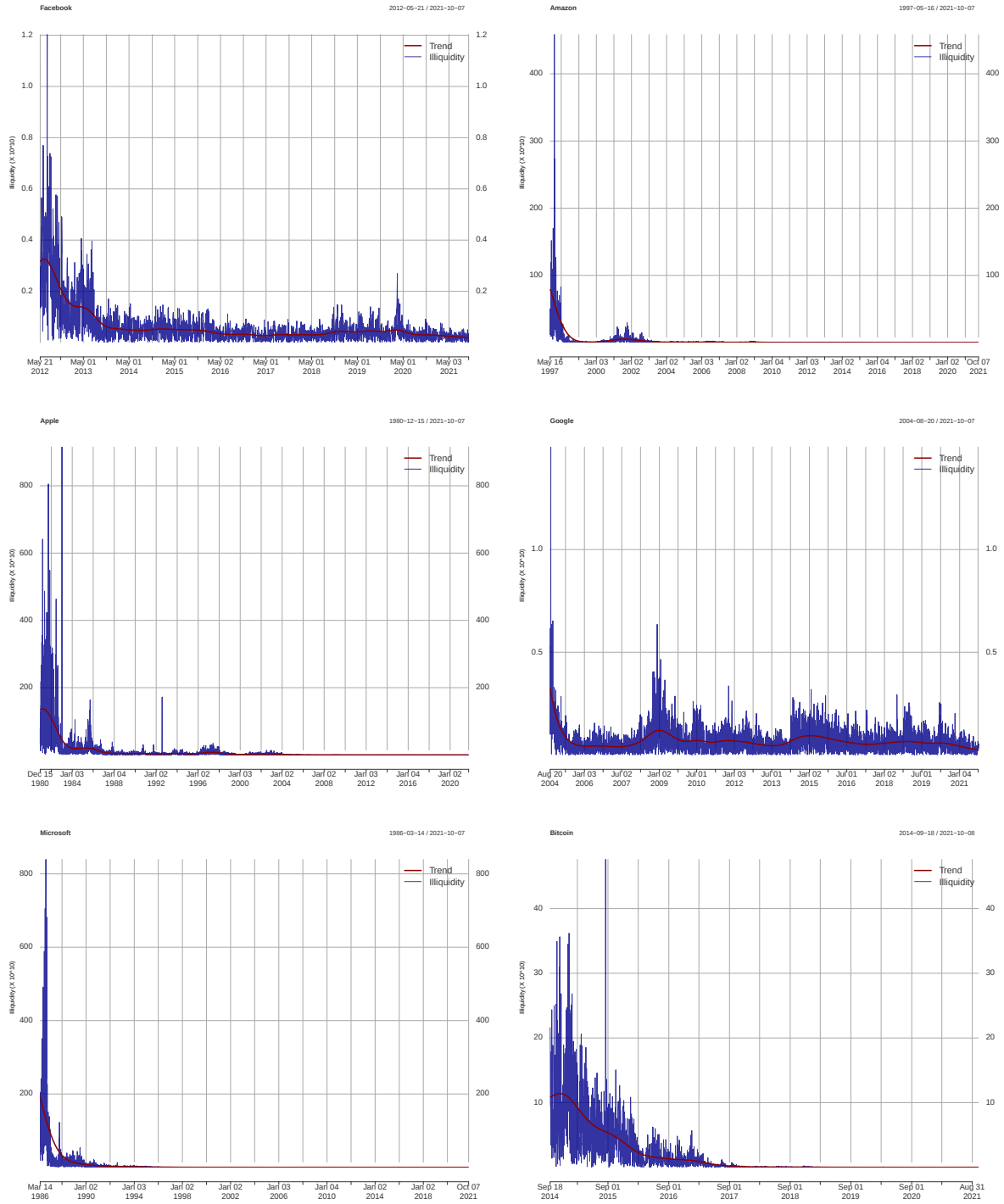


Figure 3: Fab 5 and Bitcoin illiquidity series and trend functions ($\times 10^{10}$).

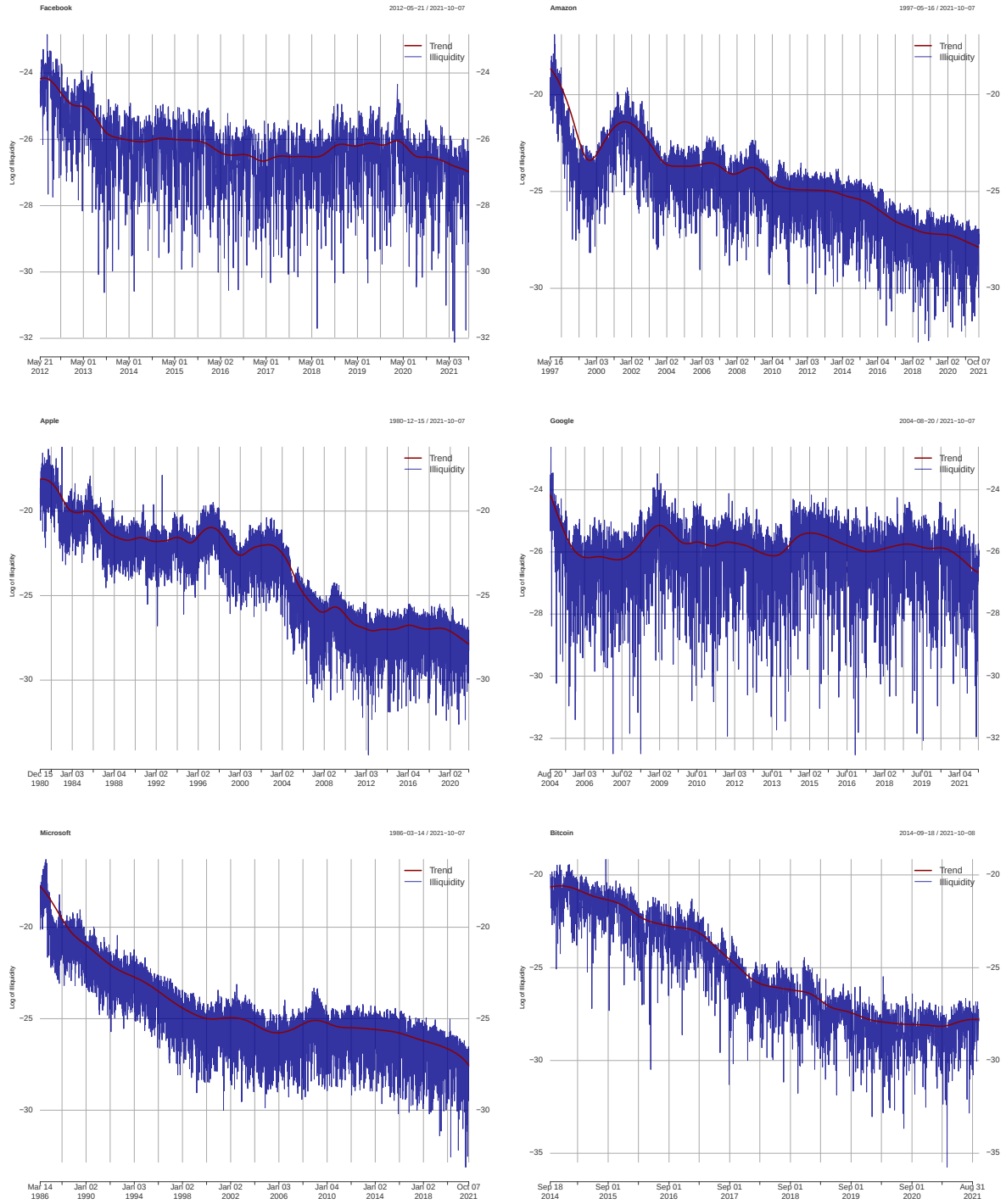


Figure 4: Fab 5 and Bitcoin log illiquidity series and trend functions.

8.2 Estimation results

We introduce the detrended illiquidity series, $\ell_t^* = \ell_t/g(t/T)$, which are assumed to be mean stationary. We then estimate the parameters θ of the λ_t process based on moment restrictions and an i.i.d. assumption for the shock distributions. We consider two model specifications for λ_t . Firstly, in the classic specification, we suppose $\lambda_t = \omega + \beta\lambda_{t-1} + \gamma\ell_{t-1}^*$ and use an expectation targeting approach to obtain ω by setting $\omega = 1 - \beta - \gamma$. Secondly, in the specification with asymmetric effect, we suppose $\lambda_t = \omega + \beta\lambda_{t-1} + \gamma\ell_{t-1}^* + \gamma^-\ell_{t-1}^*\mathbf{I}_{R_{t-1}<0}$ where R_t is the return at time t . We further assume that conditional on $\mathcal{F}_{t-1}$, R_t has a zero median and is uncorrelated with ℓ_{t-1}^* . Therefore, it implies that $E[\ell_{t-1}^*\mathbf{I}_{R_{t-1}<0} | \mathcal{F}_{t-1}] = \lambda_t/2$ and we can also use a targeting approach for ω by setting $\omega = 1 - \beta - \gamma - \gamma^-/2$.

8.2.1 Estimation based on conditional moment restrictions

We use the GMM approach based on the conditional moment restrictions to acquire an initial consistent estimators 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. $E\left[\frac{\ell_t^*}{\lambda_t} - 1 | \mathcal{F}_{t-1}\right] = 0$. We further improve the estimates of the $g(t/T)$ function using the estimated $\hat{\lambda}_t = \hat{\lambda}_t(\hat{\theta}_{GMM})$ obtained in the previous step. This, in turn, allows us to further improve the estimates of the θ parameters.

We report the obtained estimates with associated t-statistics in [Table 3](#). It can be observed that the parameter estimates in the classic specification are almost always statistically significant at the 5% level. However, the γ and γ^- estimates in the asymmetric model specification are in general not significant. The overall lack of statistical significance indicates that the asymmetric effect does not contribute to improving the empirical fit of the model based on the first moment restriction. We will further investigate whether including an asymmetric term is beneficial in the case where the models are estimated via the MLE approach under an i.i.d. shock assumption. Finally, the coefficient β is close to one, indicating high persistence in the short-run dynamics of the illiquidity series.

We improve the estimates of the trend function based on the estimated $\hat{\lambda}_t$ process. The log transforms of the initial and updated estimates of the trend function, i.e. $\log g(t/T)$, are plotted in Figure 7 of Appendix E.1 together with the log illiquidity series. We observe that the updated trend function estimates – under both the classic model and the asymmetric specification – are different from the initial estimate but only to a minor extent. 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 3: Estimated parameters of the λ_t process based on first moment restriction.

	Classic		Asymmetric		
	β	γ	β	γ	$\gamma^{(-)}$
Facebook	0.951 (20.38)	0.029 (2.06)	0.951 (27.33)	0.029 (1.21)	-0.001 (-0.04)
Amazon	0.946 (71.35)	0.052 (4.47)	0.948 (59.21)	0.042 (1.52)	0.014 (0.47)
Apple	0.914 (60.56)	0.077 (6.75)	0.914 (69.36)	0.059 (2.34)	0.029 (0.84)
Google	0.967 (64.94)	0.028 (3.08)	0.964 (71.51)	0.021 (1.95)	0.010 (1.21)
Microsoft	0.945 (88.26)	0.051 (5.95)	0.859 (9.74)	0.016 (1.50)	0.142 (1.69)
Bitcoin	0.960 (74.04)	0.036 (3.56)	0.960 (21.68)	0.021 (0.19)	0.025 (0.12)

Note: The estimated parameters are $\theta = (\beta, \gamma)$ for the classic specification and $\theta = (\beta, \gamma, \gamma^{(-)})$ for the asymmetric specification of λ_t . The numbers in parentheses are the t-statistics of the corresponding parameter estimates.

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

We estimate the model using an alternative approach – the semiparametric MLE approach – where we assume i.i.d. error terms. The conditional distribution of the error term ζ_t can be freely chosen within the class of distributions satisfying the required characteristics, namely the density having non-negative support with unit mean and variance σ_ζ^2 . In this application, we assume that the error term follows a Weibull($\Gamma(1 + \varphi)^{-1}, \varphi$) distribution, with the parameter φ controlling the shape of the distribution.⁵ Based on the local linear estimator of the $g(t/T)$ function, we first obtain a consistent estimator of the λ_t process parameters via the Quasi-Maximum Likelihood (QML) estimation approach. This is achieved by maximizing the log likelihood assuming $g(t/T)$ is known and ζ_t is parametrically specified with a Weibull density. The fully efficient estimates can then be obtained by using a one-step update approach using the efficient scores based on the initial consistent estimators as introduced in Section 4.2.1.

We report the estimated parameters with the corresponding t-statistics in Table 4. The estimates for the parameters of the λ_t process are significant and all illiquidity series exhibit a high degree of persistence as the β coefficients are close to one. In addition, the estimated shape parameter of the Weibull distribution for the error terms are ranging from 1.14 to 1.40, indicating that the volatility of ζ_t is ranging from 0.73 to 0.88. This is consistent with the observation that the five tech stocks have comparable volatility levels while the Bitcoin asset has much higher volatility.

Furthermore, we provide diagnostics on the validity of our assumptions for the error term ζ_t . Concerning the i.i.d. assumption, we plot the autocorrelation function (ACF) of ζ_t in Figure 8 and Figure 9 of Appendix E.2 respectively for the classic and asymmetric model specifications of λ_t . Similarly, we plot the ACF of ζ_t^2 under the two specifications in Figure 10 and Figure 11 of Appendix E.2. We observe that in most of the cases, there is no evidence suggesting autocorrelation in the residual or squared residual series.

⁵We consider Exponential, Gamma, and Weibull distributions in our analysis as they are commonly adopted in the literature. To save space, we focus on the Weibull distribution in the paper as it provides slightly better performance in terms of fit compared to the other two alternative distributions.

Table 4: Fully efficient estimates of the parameters for the λ_t process under the assumption that the error term ζ_t follows a Weibull distribution.

	Classic				Asymmetric				
	β	γ	φ	σ	β	γ	$\gamma^{(-)}$	φ	σ
Facebook	0.861 (5.98)	0.046 (4.94)	1.378 (804.81)	0.734	0.921 (33.88)	-0.028 (-5.77)	0.106 (9.06)	1.398 (720.95)	0.725
Amazon	0.916 (285.12)	0.080 (28.00)	1.367 (2204.62)	0.740	0.928 (346.05)	0.038 (13.18)	0.065 (21.72)	1.379 (2131.88)	0.734
Apple	0.876 (322.41)	0.095 (68.19)	1.309 (3151.33)	0.771	0.891 (360.82)	0.045 (49.13)	0.077 (27.71)	1.310 (2434.88)	0.770
Google	0.908 (31.83)	0.046 (8.18)	1.284 (5624.95)	0.785	0.932 (70.03)	-0.011 (-3.21)	0.088 (12.27)	1.302 (5351.82)	0.775
Microsoft	0.924 (361.55)	0.067 (33.24)	1.366 (1332.79)	0.740	0.927 (360.10)	0.038 (17.58)	0.052 (20.17)	1.370 (1217.23)	0.739
Bitcoin	0.892 (38.07)	0.061 (8.49)	1.137 (4774.77)	0.881	0.892 (36.48)	0.039 (6.64)	0.041 (7.15)	1.143 (4895.14)	0.877

Note: The estimated parameters are $\theta = (\beta, \gamma, \varphi)$ for the classic specification and $\theta = (\beta, \gamma, \gamma^{(-)}, \varphi)$ for the asymmetric specification of λ_t . φ is the shape parameter of the Weibull distribution which has mean 1 and standard deviation σ_ζ

of $\sqrt{\frac{\Gamma(1+\frac{2}{\varphi})}{\Gamma^2(1+\frac{1}{\varphi})}} - 1$. The numbers in parentheses are the t-statistics of the corresponding parameter estimates.

Furthermore, we use the probability integral transformation (PIT) to check how well the assumed Weibull conditional distribution fits the data. The histogram plots of the PITs shown in [Figure 12](#) and [Figure 13](#) of [Appendix E.2](#) are quite close to a uniform distribution. All assets exhibit a common pattern where the error term has ticker tail on the left-hand side and thinner tail on the right-hand side compared to a Weibull distribution. This issue can be addressed using another more flexible distribution, for example the generalized Pareto distribution. However, we do not pursue this direction in our analysis and focus instead on investigating whether using a nonparametric density help improve the model fit.

We further improve the estimation of $g(t/T)$ by maximizing the local likelihood based on the estimated $\hat{\lambda}_t$ process 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 14](#) of [Appendix E.2](#) together with the log illiquidity series. As in the GMM case (see [Section 8.2.1](#)), we observe that the updated trend function estimates are different from the initial estimate but only to a minor extent.

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

We consider whether replacing the parametric assumption for the error density f with a nonparametric kernel estimator can further improve the fit of our model to empirical data. We plot in [Figure 15](#) and [Figure 16](#) of [Appendix E.3](#) the estimated nonparametric density against the Weibull density using the shape parameter estimates from [Section 8.2.2](#). We observe that the estimated nonparametric density curves do not fall between the two standard deviation bands of the estimated Weibull densities, suggesting that the difference between the estimated parametric and nonparametric densities are statistically significant.

The estimated nonparametric density allows us to further improve the maximum likelihood estimation results for the λ_t process. We can obtain the fully efficient estimates in the nonparametric density case using the one-step update approach based on the efficient scores introduced in [Section 4.2.2](#). The estimates with associated t-statistics are reported in [Table 5](#) and the parameters are all statistically significant. Comparing the estimated values for the λ_t parameters reported in [Table 4](#) and [Table 5](#), we observe that the difference in the estimated parameter values between the parametric and nonparametric cases are overall quite small. This indicates that the QML estimation approach, combined with the one-step update based on the efficient scores to improve efficiency, provides rather accurate parameter estimates.

We further present the log likelihood computed using the parameter estimates obtained in the parametric and nonparametric cases in [Table 6](#). We conclude that the ML estimation approach assuming a Weibull distribution for the error term provides good

estimation performance, but using a nonparametric estimator for the error density can further improve performance in terms of the likelihood (the only exception is Microsoft).

Table 5: Fully efficient estimates of the parameters for λ_t process when using the non-parametric estimates of the density of the error term ζ_t .

	Classic		Asymmetric		
	β	γ	β	γ	$\gamma^{(-)}$
Facebook	0.858 (7.71)	0.049 (5.39)	0.904 (104.54)	-0.019 (-5.42)	0.102 (11.51)
Amazon	0.916 (294.21)	0.079 (24.97)	0.928 (384.91)	0.039 (14.50)	0.058 (20.19)
Apple	0.895 (175.42)	0.083 (26.28)	0.907 (238.28)	0.039 (14.98)	0.067 (19.60)
Google	0.911 (49.55)	0.046 (9.07)	0.924 (210.14)	-0.007 (-3.72)	0.086 (15.67)
Microsoft	0.929 (425.62)	0.066 (33.04)	0.931 (504.54)	0.042 (22.95)	0.044 (18.30)
Bitcoin	0.902 (66.59)	0.064 (11.04)	0.900 (67.85)	0.050 (8.93)	0.025 (3.78)

Note: The estimated parameters are $\theta = (\beta, \gamma)$ for the classic specification and $\theta = (\beta, \gamma, \gamma^{(-)})$ for the asymmetric specification of λ_t . The numbers in parentheses are the t-statistics of the corresponding parameter estimates.

Table 6: Log-likelihood comparison between models using the parametric (Weibull) and nonparametric estimates of the ζ_t density.

	Weibull		Nonparametric		Difference	
	classic	asymmetric	classic	asymmetric	classic	asymmetric
Facebook	-2171.12	-2138.88	-2148.38	-2106.34	22.74	32.55
Amazon	-4947.87	-4898.92	-4908.41	-4848.90	39.46	50.01
Apple	-8860.69	-8790.35	-8664.33	-8579.91	196.36	210.44
Google	-4020.91	-3969.86	-3969.77	-3892.27	51.14	77.59
Microsoft	-7440.48	-7398.32	-7455.04	-7404.63	-14.56	-6.32
Bitcoin	-2397.82	-2389.82	-2384.46	-2379.91	13.36	9.91

Note: The numbers reported are in terms of $\log LL$. The difference is computed as $\log LL$ in nonparametric density case minus $\log LL$ in the parametric Weibull density case.

8.3 Testing for permanent shifts: discontinuity in g function

We test for a potential discontinuity at a given time u_0 by estimating the $g^\pm(u_0)$ functions via the local linear approach. We then construct the test statistics $\tau(u_0)$ to detect whether there is a permanent shift in the illiquidity level occurring at time u_0 . To facilitate the computation of the asymptotic variance of $\hat{g}^\pm(u_0)$, we work with the improved estimator obtained by smoothing out ℓ_t , i.e. $\ell_t/\hat{\lambda}_t$. We plot in [Figure 5](#) and [Figure 6](#) the test statistics $\tau(u_0)$ for the five tech stocks and the Bitcoin asset over the corresponding sample period.

We start with a typical stock specific-event, a stock split, and test for permanent shifts in the liquidity dynamics arising after stock splits. The five tech stocks we consider have quite different corporate policies regarding shareholders and in particular their propensity to split their stock differs. In our study, Facebook never split its stock, Amazon split its stock three times but the last time being before 2000 (perhaps coincidentally these were all in the pre-decimal era). Microsoft split its stock 9 times in our sample period but the last time was in 2002. Google split its stock twice in our sample in 2014 and 2015, but not before that or since. Apple is a regular splitter with 5 splits in our sample fairly evenly spaced in time. Each split is marked as a red dot on the curves in [Figure 5](#) and [Figure 6](#). The majority of the statistics on stock split dates is outside of the 5% critical value bands, suggesting an overall significance of the stock split events.

In addition, [Table 7](#) provides the average test statistics for each stock on their stock split dates together with the average across all stock split events for the four considered stocks. Firstly, we should note that the average test statistic τ is positive in all cases, indicating an increase in stock illiquidity and thus a corresponding decrease in stock liquidity after the splits. Secondly, we observe that the average statistic indicates a significant difference between pre- and post-split long-term trends of the illiquidity series. This suggests that the decrease in liquidity after stock splits is permanent and significant.

To test for temporary effects of stock splits on the liquidity level, we need to normalize the illiquidity series using the estimated one-sided trend functions $\hat{g}^\pm(u)$. Once the detrended illiquidity series are obtained, we use the consistent test developed in Section

Table 7: Average statistics for testing permanent breaks in the liquidity series.

	Amazon	Apple	Google	Microsoft	Average
τ	6.481	5.499	5.404	4.423	5.134

6 to test the null hypothesis that $\alpha_1 = \alpha_2 = \dots = \alpha_J = 0$ against the general alternative semiparametric model with the assumption that the error terms follow a Weibull distribution. Here, we consider a five-day window, namely from two days before until two days after the stock split date. We report in Table 8 and Table 9 the test statistic values for the permanent (τ_{LR}) and temporary (τ_{SR}) shifts together with the 2.5% and 97.5% quantiles of τ_{SR} which are estimated based on past data.⁶ We observe that the effect of stock splits on the short-term dynamics of liquidity is almost always not significant. The only two exceptions are for Microsoft stock splits in 1991 and 1996 which were associated with a significant short-term shift in liquidity dynamics. Therefore, our empirical evidence suggest overall that stock splits of tech companies had a significant permanent effect on the long-run trend of their illiquidity process but not on the short-run dynamics.

⁶Note that for the split events preceded by another one we only consider the period after the first stock split event for the computation of the τ_{SR} quantiles.

Table 8: Test statistics for detecting permanent and temporary breaks in the liquidity series of Amazon and Apple stocks.

		Split 1	Split 2	Split 3		
		Date	1998-06-02	1999-01-05	1999-09-02	
Amazon	Splits	02:01	03:01	02:01		
	τ_{LR}	2.58	16.72	0.15		
	τ_{SR}	19.92	4.49	-17.14		
	$Q_{SR}^{2.5\%}$	-29.38	-30.55	-32.61		
	$Q_{SR}^{97.5\%}$	47.28	33.40	6.88		
Apple		Split 1	Split 2	Split 3	Split 4	Split 5
	Date	1987-06-16	2000-06-21	2005-02-28	2014-06-09	2020-08-31
	Splits	02:01	02:01	02:01	07:01	04:01
	τ_{LR}	7.03	3.29	12.45	1.64	3.09
	τ_{SR}	0.18	8.51	-2.94	3.10	8.59
	$Q_{SR}^{2.5\%}$	-14.78	-12.81	-10.77	-11.19	-9.74
	$Q_{SR}^{97.5\%}$	18.50	15.31	16.16	10.63	11.99

Note: We report the test statistic values for the permanent (τ_{LR}) and temporary (τ_{SR}) shifts together with the 2.5% and 97.5% quantiles of τ_{SR} which are estimated based on past data.

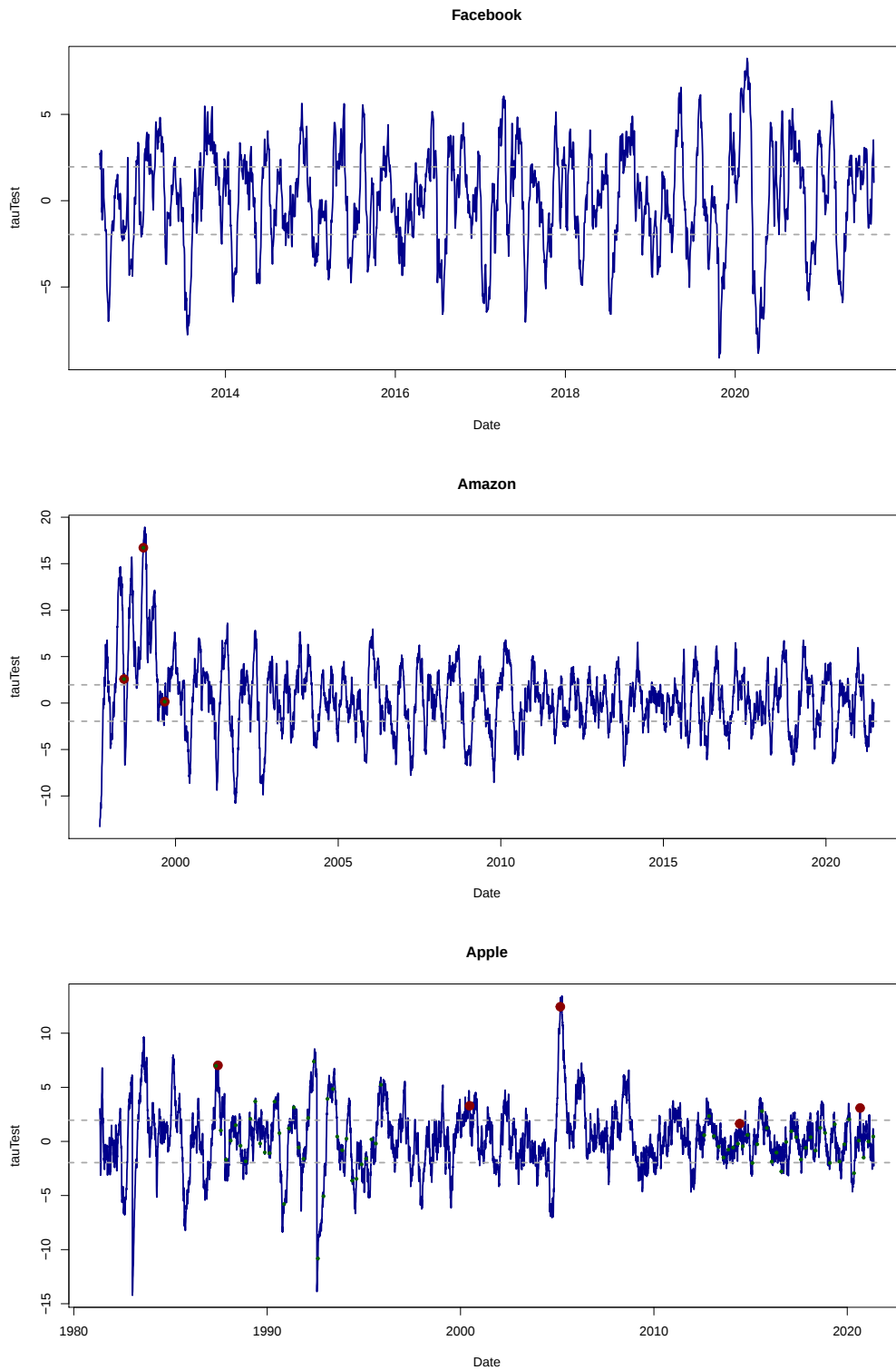


Figure 5: Test statistics for detecting permanent breaks in the illiquidity series.

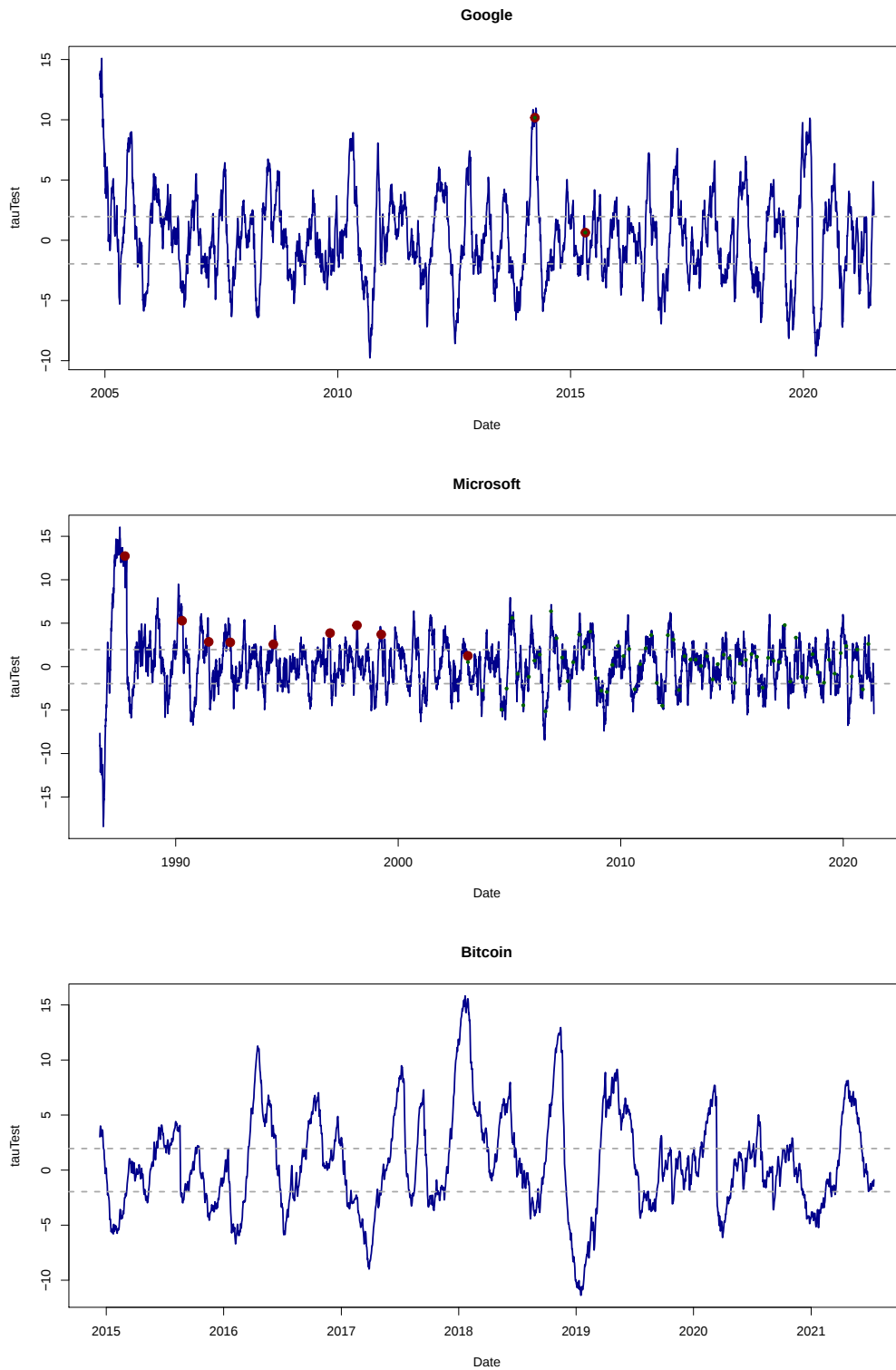


Figure 6: Test statistics for detecting permanent breaks in the illiquidity series.

Table 9: Test statistics for detecting permanent and temporary breaks in the liquidity series of Google and Microsoft stocks.

Google	Split 1		Split 2			
	Date	2014-03-27	2015-04-27			
	Splits	2002:1000	10027455:10000000			
	τ_{LR}	10.17	0.64			
	τ_{SR}	-2.94	8.79			
	$Q_{SR}^{2.5\%}$	-20.58	-14.52			
	$Q_{SR}^{97.5\%}$	20.76	22.63			
Microsoft	Split 1		Split 2	Split 3	Split 4	Split 5
	Date	1987-09-21	1990-04-16	1991-06-27	1992-06-15	1994-05-23
	Splits	02:01	02:01	03:02	03:02	02:01
	τ_{LR}	12.73	5.28	2.86	2.80	2.57
	τ_{SR}	8.58	-0.54	31.94	6.65	-6.29
	$Q_{SR}^{2.5\%}$	-30.60	-20.84	-16.34	-17.61	-17.06
	$Q_{SR}^{97.5\%}$	24.59	20.14	26.16	17.41	20.70
	Split 6		Split 7	Split 8	Split 9	
	Date	1996-12-09	1998-02-23	1999-03-29	2003-02-18	
	Splits	02:01	02:01	02:01	02:01	
	τ_{LR}	3.86	4.76	3.71	1.25	
	τ_{SR}	24.00	4.85	9.10	12.17	
	$Q_{SR}^{2.5\%}$	-15.11	-16.06	-17.81	-19.56	
	$Q_{SR}^{97.5\%}$	17.04	17.55	16.24	25.43	

Note: We report the test statistic values for the permanent (τ_{LR}) and temporary (τ_{SR}) shifts together with the 2.5% and 97.5% quantiles of τ_{SR} which are estimated based on past data.

8.4 Risk premium

Amihud (2002) studies how illiquidity, captured by his illiquidity measure A_t introduced in Section 2, relates to stock excess returns in both the time series and cross-sectional dimensions. We build on this analysis to investigate the effect of each component of the S&P 500 index illiquidity process – i.e. the expected long-term and short-term components (respectively $g(t/T)$ and λ_t) and illiquidity shocks (ζ_t) – on the stock market index excess returns (the market “risk premium”). We consider three frequencies in our analysis – daily, weekly and monthly. The S&P 500 index illiquidity and log illiquidity series together with the stock market index return data for the three considered frequencies are plotted respectively in Figure 18, Figure 19 and Figure 20 of Appendix E.4. We note that there exists a strong downward trend in the illiquidity process while the return series is somewhat stationary. This suggests that the relationship between the long-run trend of market liquidity and the stock excess return would be less significant.⁷ Therefore, we focus on detrended illiquidity and market excess return series to study the effect of expected short-run illiquidity variations and unexpected illiquidity shocks on the market risk premium.

We consider the specification from Equation (31) in Section 7 for the regression of detrended risk premium on illiquidity components.⁸ The estimation results for the three sampling frequencies considered are provided in Table 10. We observe that the estimated γ coefficients for the short-run expected illiquidity component λ_t are positive and significant which indicates that the expected market excess return is an increasing function of the short-run expected illiquidity process. This observation is consistent with the intuition that higher expected market illiquidity would make investors demand higher excess returns on stocks as a compensation for gaining exposure to this source of risk. Moreover, the estimated δ coefficients for the shock term ζ_t are negative and significant, suggesting that

⁷This is confirmed by regression results based on Equation (30) introduced in Section 7. The coefficient estimates for the parameter b associated with the long-run trend illiquidity component are not significant. Results are available from the authors upon request.

⁸The time-varying unconditional equity premium $m(t/T)$ is obtained via a local linear estimator.

the unexpected market illiquidity has a negative effect on the stock excess return. This can be explained by the fact that stock prices would likely fall when illiquidity unexpectedly rises, thus decreasing expected returns.

Table 10: Coefficient estimates for regressions using daily, weekly and monthly observations.

	Daily	Weekly	Monthly
α	-0.0007 (0.0005)	-0.0044 (0.0032)	-0.0513** (0.0176)
γ	0.0017*** (0.0004)	0.0066* (0.0032)	0.0561** (0.0177)
δ	-0.0010*** (0.0001)	-0.0023*** (0.0006)	-0.0045* (0.0022)
R^2	0.0068	0.0073	0.0187
Adj. R^2	0.0066	0.0066	0.0161
Num. obs.	15387	2893	741

Note: We estimate the regression based on Equation (31): $R_{mt} - R_{ft} - m(t/T) = \alpha + \gamma \times \lambda_t + \delta \times \zeta_t + \epsilon_t$, where $m(t/T)$ is the time-varying unconditional equity premium. The significance level is indicated by *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

9 Conclusions

The motivation for this paper stems from the observation that financial market illiquidity dynamics across various asset classes are driven by both low-frequency and higher-frequency variations, which makes the stationarity assumption unreasonable for illiquidity modelling. We develop a class of a dynamic autoregressive models that captures the slow-varying long-term trend with a nonparametric component and the short-run variations in the illiquidity series with an autoregressive component. We provide estimation theory for the GMM estimator and the efficient semiparametric ML estimator for the parametric and nonparametric error density cases. An empirical application – using the five largest US technology stocks and the Bitcoin asset – demonstrates the good performance of our framework in capturing the salient features of illiquidity dynamics.

We further develop a methodology to detect the occurrence of permanent and temporary shifts in the illiquidity process at a given point in time. We apply this framework to study how stock splits affect liquidity dynamics. Clearly stock splits are only one of many events that seem to permanently shift the stock price, quarterly earnings announcements, new product releases, and macroeconomic news are all known to have big effects on the prices and trading volumes of these stocks in particular. Nevertheless, we do find a significant negative effect of stock splits on the long-run trend level of liquidity around the time of the stock splits themselves, while the effect on the short-run illiquidity dynamics is not significant. Our results are broadly consistent with [Copeland \(1979\)](#).

Lastly, we investigate the link between stock market excess returns and the different components of illiquidity for the S&P 500 stock market index. We show that, while excess returns are an increasing function of the expected illiquidity component, unexpected illiquidity shocks decrease stock prices and returns. Our finding is consistent with the findings of [Amihud \(2002\)](#) based on his cruder methodology.

Appendices

A Lemmas

Lemma 1. *We have $\widehat{g}(u) - g(u) = V_T(u) + B_T(u)$, where $B_T(u)$ is deterministic and*

$$\sup_{u \in [0,1]} |V_T(u)| = O_P \left(\sqrt{\frac{\log T}{Th}} \right), \quad \sup_{u \in [0,1]} |B_T(u)| = O(h^2).$$

Proof of Lemma 1. We write

$$\widehat{g}(u) - g(u) = \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) v_t + \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) g(t/T) - g(u) = V_T(u) + B_T(u), \quad (32)$$

where B_T is a deterministic bias term and V_T is a mean zero stochastic term. The result follows by standard arguments.

B Proof of main results

Proof of Theorem 1. From the expansion in Equation (32), we have $V_T(u) = \sum_{t=1}^T K_h(t/T - u) v_t / T$, and we may show that

$$\sqrt{Th} V_T(u) \implies N(0, \|K\|_2^2 g(u)^2 \sigma_v^2),$$

where σ_v^2 is the long run variance.

Proof of Theorem 2. First, note that

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{\ell_t}{\lambda_t} - g(u) &= \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) g(t/T) \zeta_t - g(u) \\ &= \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) g(t/T) (\zeta_t - 1) \\ &\quad + \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) g(t/T) - g(u) \\ &= V_T^+(u) + B_T^+(u), \end{aligned}$$

where $V_T^+(u)$ is a mean zero stochastic term, whereas $B_T^+(u) = B_T(u)$ is the deterministic bias term. The term $V_T^+(u)$ has a MDS error term and satisfies the CLT

$$\sqrt{Th}V_T^+(u) \implies N\left(0, \|K\|_2^2 g(u)^2 \sigma_\zeta^2\right).$$

We next show that this is the leading term.

We have

$$\begin{aligned} \left| \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{\ell_t}{\hat{\lambda}_t} - \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{\ell_t}{\lambda_t} \right| &= \left| -\frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{\ell_t}{\lambda_t} \frac{\lambda_t(\hat{\theta}, \hat{g}) - \lambda_t}{\lambda_t(\hat{\theta}, \hat{g})} \right| \\ &\leq \max_{1 \leq t \leq T} \left| \frac{1}{\lambda_t(\hat{\theta}, \hat{g})} \right| \left| \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{\ell_t}{\lambda_t} (\lambda_t(\hat{\theta}, \hat{g}) - \lambda_t) \right| \\ &\leq O_P(1) \times \left| \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{\ell_t}{\lambda_t} (\lambda_t(\hat{\theta}, \hat{g}) - \lambda_t) \right| \\ &= o_P(T^{-1/2} h^{1/2}), \end{aligned}$$

because

$$\begin{aligned} \lambda_t(\hat{\theta}, \hat{g}) &= \lambda_t(\theta_0, g_0) - \left| \lambda_t(\hat{\theta}, \hat{g}) - \lambda_t(\theta_0, g_0) \right| \\ &\geq \lambda_t(\theta_0, g_0) - o_P(1) \end{aligned}$$

by the triangle inequality and the uniform convergence of $\hat{g}$ given in Lemma 1.

Proof of Theorem 3. We apply [Chen et al. \(2003\)](#). The key thing is to determine the contribution of the nonparametric estimation to the variance of the score function.

We have

$$M_T(\theta, \hat{g}) = \frac{1}{T} \sum_{t=1}^T \rho_t(\theta, g_0) + \frac{1}{T} \sum_{t=1}^T z_{t-1} (\hat{\ell}_t^* - \ell_t^* - (\hat{\lambda}_t(\theta) - \lambda_t(\theta))).$$

We consider

$$M_T(\theta, g) = \frac{1}{T} \sum_{t=1}^T \rho_t(\theta, g), \quad \rho_t(\theta, g) = z_{t-1} \left(\frac{\ell_t}{g(t/T)} - \lambda_t(\theta, g) \right)$$

$$\rho_t(\theta_0, g_0) = z_{t-1} \lambda_t (\zeta_t - 1)$$

$$\lambda_t(\theta, g) = 1 - \beta - \gamma + \beta \lambda_{t-1} + \gamma \frac{\ell_{t-1}}{g((t-1)/T)} = \frac{1 - \beta - \gamma}{1 - \beta} + \gamma \sum_{j=1}^{\infty} \beta^{j-1} \frac{\ell_{t-j}}{g((t-j)/T)}.$$

We have

$$M(\theta, g) = \lim_{T \rightarrow \infty} E(M_T(\theta, g)).$$

We define

$$\Gamma_2(\theta, g_0) \circ (g - g_0) = \frac{\partial}{\partial \tau} M(\theta, g_0 + \tau(g - g_0)).$$

We have

$$\frac{\lambda_t(\theta, g_0 + \tau(g - g_0)) - \lambda_t(\theta, g_0)}{\tau} \simeq -\gamma \sum_{j=1}^{\infty} \beta^{j-1} \frac{\ell_{t-j}}{g_0((t-j)/T)} \frac{g((t-j)/T) - g_0((t-j)/T)}{g_0((t-j)/T)}$$

and so

$$\begin{aligned} \lim_{\tau \rightarrow 0} E \left[\frac{\lambda_t(\theta, g_0 + \tau(g - g_0)) - \lambda_t(\theta, g_0)}{\tau} \right] &= -\gamma \sum_{j=1}^{\infty} \beta^{j-1} \frac{g((t-j)/T) - g_0((t-j)/T)}{g_0((t-j)/T)} \\ &\simeq -\frac{g(t/T) - g_0(t/T)}{g_0(t/T)} \frac{\gamma}{1 - \beta} \end{aligned}$$

Furthermore,

$$\begin{aligned} \frac{\frac{\ell_t}{g(t/T)} - \frac{\ell_t}{g_0(t/T)}}{\tau} &\simeq -\frac{\ell_t}{g_0(t/T)} \frac{g(t/T) - g_0(t/T)}{g_0(t/T)} \\ E \left(\frac{\frac{\ell_t}{g(t/T)} - \frac{\ell_t}{g_0(t/T)}}{\tau} \right) &\simeq -\frac{g(t/T) - g_0(t/T)}{g_0(t/T)} \end{aligned}$$

Therefore,

$$\begin{aligned} M_T(\theta, \hat{g}) &= M_T(\theta, g_0) + \Gamma_2(\theta_0, g_0) \circ (\hat{g} - g_0) \\ &= \frac{1}{T} \sum_{t=1}^T \left(\rho_t(\theta_0, g_0) + \frac{1 - \beta - \gamma}{1 - \beta} z_{t-1} \frac{\hat{g}(t/T) - g_0(t/T)}{g_0(t/T)} \right). \end{aligned}$$

We have

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T z_{t-1} \frac{\hat{g}(t/T) - g_0(t/T)}{g_0(t/T)} &= \frac{1}{T} \sum_{t=1}^T z_{t-1} \frac{1}{T} \sum_{s=1}^T K_h(s/T - t/T) (\lambda_s \zeta_s - 1) + O(h^2) \\ &= \frac{1}{T} \sum_{s=1}^T (\lambda_s \zeta_s - 1) \frac{1}{T} \sum_{t=1}^T z_{t-1} K_h(s/T - t/T) \\ &\simeq \frac{1}{T} \sum_{s=1}^T (\lambda_s \zeta_s - 1) E(z_{s-1}). \end{aligned}$$

It follows that

$$M_T(\theta, \widehat{g}) = \frac{1}{T} \sum_{t=1}^T \left(\lambda_t(\zeta_t - 1)z_{t-1} + \frac{1 - \beta - \gamma}{1 - \beta}(\lambda_t \zeta_t - 1)E(z_{t-1}) \right) + o_P(T^{-1/2}).$$

Proof of Theorem 4. Let $\ell_t^{**} = \ell_t / \lambda_t$ then

$$\ell_t^{**} = g(t/T)\zeta_t.$$

The local likelihood is apart from a constant

$$L(g; u) = \sum_{t=1}^T K_h(t/T - u) \left(-\log g + \log f \left(\frac{\ell_t^{**}}{g} \right) \right).$$

Let $s_1(\zeta) = -f'_0(\zeta)/f_0(\zeta)$ and $s_2(\zeta) = -((f'_0(\zeta)/f_0(\zeta))\zeta + 1)$. We have

$$\frac{\partial L(g; u)}{\partial g} = \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \left(-\frac{1}{g} - \frac{1}{g} \frac{f' \left(\frac{\ell_t^{**}}{g} \right)}{f \left(\frac{\ell_t^{**}}{g} \right)} \frac{\ell_t^{**}}{g} \right) = \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{1}{g} s_2 \left(\frac{\ell_t^{**}}{g} \right)$$

$$\frac{\partial^2 L(g; u)}{\partial g^2} = \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \left(\frac{-1}{g^2} s_2 \left(\frac{\ell_t^{**}}{g} \right) - \frac{1}{g^2} s'_2 \left(\frac{\ell_t^{**}}{g} \right) \frac{\ell_t^{**}}{g} \right)$$

We have

$$\begin{aligned} \frac{\partial L(g_0(u); u)}{\partial g} &= \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{1}{g(u)} s_2(\zeta_t) \\ \frac{\partial^2 L(g_0(u); u)}{\partial g^2} &\simeq \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{1}{g(u)^2} s'_2(\zeta_t) \zeta_t. \end{aligned}$$

We have $s_2(\zeta) = 1 + \zeta f'(\zeta) / f(\zeta)$

$$E(s'_2(\zeta_t)) = \int s'_2(\zeta) \zeta f(\zeta) d\zeta = - \int s_2(\zeta) \zeta f'(\zeta) d\zeta - \int s_2(\zeta) f(\zeta) d\zeta = - \int s_2^2(\zeta) f(\zeta) d\zeta = -I_2(f).$$

C Semiparametric efficiency

C.1 Known f

Suppose that

$$\ell_t = g_\delta(t/T) \lambda_t(\theta) \zeta_t$$

$$\lambda_t = 1 - \beta - \gamma + \beta\lambda_{t-1} + \gamma\lambda_{t-1}\zeta_{t-1}$$

where ζ_t is i.i.d. with mean one and density f supported on $\mathbb{R}_+$, so that $E(\lambda_t) = 1$ and $E(\zeta_t) = 1$. We suppose that g is unknown but we consider the parameterization by δ . We first suppose that f is known. Let $s_2(\zeta) = -(1 + \zeta f'(\zeta)/f(\zeta))$.

Consider the log likelihood

$$L(\theta, \delta | \ell_1, \dots, \ell_T) = - \sum_{t=1}^T \log \lambda_t(\theta, \delta) - \sum_{t=1}^T \log g_\delta(t/T) + \sum_{t=1}^T \log f(\zeta_t(\theta, \delta))$$

$$\lambda_t(\theta, \delta) = 1 - \beta - \gamma + \beta\lambda_{t-1}(\theta, \delta) + \gamma \frac{\ell_{t-1}}{g_\delta((t-1)/T)}, \quad (33)$$

$$\zeta_t(\theta, \delta) = \frac{\ell_t}{\lambda_t(\theta, \delta)g_\delta(t/T)}. \quad (34)$$

Note that λ_t depends implicitly on δ .

We have (at the true values)

$$\frac{\partial \zeta_t(\theta, \delta)}{\partial \theta} = \frac{-\ell_t}{\lambda_t(\theta, \delta)g_\delta(t/T)} \frac{\partial \log \lambda_t}{\partial \theta} = -\zeta_t \frac{\partial \log \lambda_t}{\partial \theta}.$$

$$\begin{aligned} \frac{\partial \zeta_t(\theta, \delta)}{\partial \delta} &= \frac{-\ell_t}{\lambda_t(\theta, \delta)g_\delta(t/T)} \left(\frac{\partial \log \lambda_t}{\partial \delta} + \frac{\partial \log g_\delta(t/T)}{\partial \delta} \right) \\ &= -\zeta_t \left(\frac{\partial \log \lambda_t}{\partial \delta} + \frac{\partial \log g_\delta(t/T)}{\partial \delta} \right). \end{aligned}$$

The score functions are

$$\frac{\partial L}{\partial \theta} = \sum_{t=1}^T \frac{f'}{f}(\zeta_t) \frac{\partial \zeta_t}{\partial \theta} - \frac{\partial \log \lambda_t}{\partial \theta} = \sum_{t=1}^T s_2(\zeta_t) \frac{\partial \log \lambda_t}{\partial \theta}.$$

Furthermore,

$$\begin{aligned} \frac{\partial \lambda_t(\theta, \delta)}{\partial \beta} &= \beta \frac{\partial \lambda_{t-1}(\theta, \delta)}{\partial \beta} + \lambda_t(\theta, \delta) - 1 \\ (1 - \beta L) \frac{\partial \lambda_t(\theta, \delta)}{\partial \beta} &= \lambda_t - 1 = \beta(\lambda_{t-1} - 1) + \gamma u_{t-1}, \quad u_{t-1} = \lambda_{t-1}\zeta_{t-1} - 1 \\ \frac{\partial \log \lambda_t(\theta, \delta)}{\partial \beta} &= \frac{(1 - \beta L)^{-1} \beta(\lambda_{t-1} - 1) + \gamma (1 - \beta L)^{-1} u_{t-1}}{\lambda_t} \\ \frac{\partial \lambda_t(\theta, \delta)}{\partial \gamma} &= \beta \frac{\partial \lambda_{t-1}(\theta, \delta)}{\partial \gamma} + u_{t-1} \end{aligned}$$

$$(1 - \beta L) \frac{\partial \lambda_t(\theta, \delta)}{\partial \gamma} = u_{t-1}$$

$$\frac{\partial \log \lambda_t(\theta, \delta)}{\partial \gamma} = \frac{(1 - \beta L)^{-1} u_{t-1}}{\lambda_t}.$$

Here, L is the lag operator. We next consider the score wrt δ ,

$$\frac{\partial \lambda_t(\theta, \delta)}{\partial \delta} = \beta \frac{\partial \lambda_{t-1}(\theta, \delta)}{\partial \delta} - \gamma \frac{\ell_{t-1}}{g_\delta((t-1)/T)} \frac{\partial g_\delta((t-1)/T)/\partial \delta}{g_\delta((t-1)/T)}.$$

Therefore,

$$(1 - \beta L) \frac{\partial \lambda_t(\theta, \delta)}{\partial \delta} = -\gamma \lambda_{t-1} \zeta_{t-1} \frac{\partial \log g_\delta((t-1)/T)}{\partial \delta}$$

$$\frac{\partial \log \lambda_t(\theta, \delta)}{\partial \delta} \simeq -\gamma \frac{\partial \log g_\delta(t/T)}{\partial \delta} \frac{(1 - \beta L)^{-1} \lambda_{t-1} \zeta_{t-1}}{\lambda_t}.$$

The latter argument follows essentially because for a summable sequence $\{\psi_j\}$ and smooth function g we have

$$\sum_{j=1}^T \psi_j g\left(\frac{t-j}{T}\right) = g\left(\frac{t}{T}\right) \sum_{j=1}^T \psi_j - \frac{1}{T} \sum_{j=1}^T \psi_j j g'\left(\frac{s^*(t, j)}{T}\right) \simeq g\left(\frac{t}{T}\right) \sum_{j=1}^T \psi_j.$$

Therefore,

$$\begin{aligned} \frac{\partial L}{\partial \delta} &= \sum_{t=1}^T \frac{f'}{f}(\zeta_t) \frac{\partial \zeta_t}{\partial \delta} - \frac{\partial \log \lambda_t}{\partial \delta} - \frac{\partial \log g_\delta(t/T)}{\partial \delta} \\ &= \sum_{t=1}^T s_2(\zeta_t) \left(\frac{\partial \log \lambda_t}{\partial \delta} + \frac{\partial \log g_\delta(t/T)}{\partial \delta} \right) \end{aligned} \tag{35}$$

$$\begin{aligned} &= \sum_{t=1}^T s_2(\zeta_t) \frac{\partial \log g_\delta(t/T)}{\partial \delta} \left(1 - \gamma \frac{(1 - \beta L)^{-1} \lambda_{t-1} \zeta_{t-1}}{\lambda_t} \right) \\ &= \frac{1 - \beta - \gamma}{1 - \beta} \sum_{t=1}^T s_2(\zeta_t) \frac{1}{\lambda_t} \frac{\partial \log g_\delta(t/T)}{\partial \delta}, \end{aligned} \tag{36}$$

since

$$\begin{aligned} 1 - \gamma \frac{(1 - \beta L)^{-1} \lambda_{t-1} \zeta_{t-1}}{\lambda_t} &= \frac{\lambda_t - \gamma (1 - \beta L)^{-1} \lambda_{t-1} \zeta_{t-1}}{\lambda_t} \\ &= \frac{\lambda_t - \frac{\gamma}{1-\beta} - \gamma (1 - \beta L)^{-1} u_{t-1}}{\lambda_t} \\ &= \frac{1}{\lambda_t} \times \frac{1 - \beta - \gamma}{1 - \beta} \end{aligned}$$

$$(1 - \beta L) \lambda_t = 1 - \beta - \gamma + \gamma \lambda_{t-1} \zeta_{t-1} = 1 - \beta + \gamma u_{t-1}$$

$$\lambda_t = 1 + \gamma (1 - \beta L)^{-1} u_{t-1}.$$

Therefore, the tangent space for g consists of functions of the form

$$\mathbb{T}_g = \left\{ \sum_{t=1}^T s_2(\zeta_t) \frac{1}{\lambda_t} h(t/T) : h \in L_2[0, 1] \right\}. \quad (37)$$

That is, the score wrt g is of the form

$$\frac{\partial L}{\partial \delta} = \sum_{t=1}^T s_2(\zeta_t) \frac{1}{\lambda_t} h(t/T)$$

for some function $h(\cdot)$. Because of the presence of $1/\lambda_t$ this is different from our volatility paper...

The efficient score function L_θ^* for θ in the presence of unknown g (but known φ) is the residual from the projection of L_θ onto the tangent space $\mathbb{T}_g$, this is

$$\begin{aligned} L_\theta^* &= \sum_{t=1}^T s_2(\zeta_t) \left(\frac{\partial \log \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \log \lambda_t}{\partial \theta} \frac{1}{\lambda_t} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \frac{1}{\lambda_t} \right) \\ &= \sum_{t=1}^T s_2(\zeta_t) \frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right). \end{aligned} \quad (38)$$

This can be verified as for any element of $\mathbb{T}_g$ (indexed by $h(\cdot)$) we have

$$\begin{aligned} &E \left[s_2(\zeta_t) \frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) \times s_2(\zeta_t) \frac{1}{\lambda_t} h(t/T) \right] \\ &= E [s_2^2(\zeta_t)] E \left[\frac{1}{\lambda_t^2} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) \right] h(t/T) \\ &= E [s_2^2(\zeta_t)] \left[E \left(\frac{1}{\lambda_t^2} \frac{\partial \lambda_t}{\partial \theta} \right) - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} E \left(\frac{1}{\lambda_t^2} \right) \right] h(t/T) \\ &= 0 \end{aligned}$$

for any $h(\cdot)$.

C.2 Parametric f

Now suppose that $f = f_\varphi$, where φ is unknown and has to be estimated. Then

$$L(\theta, \varphi, \delta | \ell_1, \dots, \ell_T) = - \sum_{t=1}^T \log \lambda_t(\theta, \delta) - \sum_{t=1}^T \log g_\delta(t/T) + \sum_{t=1}^T \log f_\varphi(\zeta_t(\theta, \delta)),$$

where f_φ is a density function that imposes through its parameterization the unit mean assumption. We have

$$\frac{\partial L}{\partial \varphi} = \sum_{t=1}^T \frac{\partial f_\varphi(\zeta_t) / \partial \varphi}{f_\varphi(\zeta_t)}.$$

We have

$$E \left(\frac{\partial L}{\partial \varphi} \frac{\partial L}{\partial \delta} \right) = \sum_{t=1}^T E \left(\frac{\partial f_\varphi(\zeta_t) / \partial \varphi}{f_\varphi(\zeta_t)} s_2(\zeta_t) \right) E \left(\frac{1}{\lambda_t} \right) h(t/T).$$

We have

$$\begin{aligned} E \left(\frac{\partial f_\varphi(\zeta_t) / \partial \varphi}{f_\varphi(\zeta_t)} s_2(\zeta_t) \right) &= \int \frac{\partial f_\varphi(\zeta) / \partial \varphi}{f_\varphi(\zeta)} s_2(\zeta) f_\varphi(\zeta) d\zeta \\ &= \int s_2(\zeta) \frac{\partial f_\varphi(\zeta)}{\partial \varphi} d\zeta \\ &= - \int \left(1 + \zeta \frac{f'_\varphi(\zeta)}{f_\varphi(\zeta)} \right) \frac{\partial f_\varphi(\zeta)}{\partial \varphi} d\zeta \\ &= - \int \zeta \left(\frac{f'_\varphi(\zeta)}{f_\varphi(\zeta)} - 1 \right) \frac{\partial f_\varphi(\zeta)}{\partial \varphi} d\zeta. \end{aligned}$$

Therefore,

$$E \left(\frac{\partial L}{\partial \varphi} \frac{\partial L}{\partial \delta} \right) = \sum_{t=1}^T E \left(\frac{\partial f_\varphi(\zeta_t) / \partial \varphi}{f_\varphi(\zeta_t)} s_2(\zeta_t) \right) E \left(\frac{1}{\lambda_t} \right) h(t/T) \neq 0$$

for any parameterization of g and so the efficient score function for φ in the presence of unknown g is

$$L_\varphi^* = \sum_{t=1}^T \left(\frac{\partial f_\varphi(\zeta_t) / \partial \varphi}{f_\varphi(\zeta_t)} - \frac{E \left(\frac{\partial f_\varphi(\zeta_t) / \partial \varphi}{f_\varphi(\zeta_t)} s_2(\zeta_t) \right) E \left(\frac{1}{\lambda_t} \right)}{I_2(f) E \left(\frac{1}{\lambda_t^2} \right)} s_2(\zeta_t) \frac{1}{\lambda_t} \right). \quad (39)$$

A special case where φ is a scale parameter, i.e.,

$$f_\varphi(\zeta) = \frac{1}{\varphi} f_0 \left(\frac{\zeta}{\varphi} \right).$$

We have

$$f'_\varphi(\zeta) = \frac{1}{\varphi^2} f'_0\left(\frac{\zeta}{\varphi}\right),$$

$$s_2(\zeta) = -\left(1 + \frac{f'_\varphi(\zeta)}{f_\varphi(\zeta)} \zeta\right) = -\left(1 + \frac{f'_0\left(\frac{\zeta}{\varphi}\right)}{f_0\left(\frac{\zeta}{\varphi}\right)} \left(\frac{\zeta}{\varphi}\right)\right) = -\left(1 + \frac{f'_0(z)}{f_0(z)} z\right),$$

where $z = \zeta/\varphi$. In that case

$$\begin{aligned} \frac{\partial f_\varphi(\zeta_t)}{\partial \varphi} &= \frac{-1}{\varphi^2} f_0\left(\frac{\zeta}{\varphi}\right) + \frac{1}{\varphi} f'_0\left(\frac{\zeta}{\varphi}\right) \frac{\zeta}{\varphi^2} \\ \frac{\partial f_\varphi(\zeta_t)/\partial \varphi}{f_\varphi(\zeta_t)} &= \frac{\frac{-1}{\varphi^2} f_0\left(\frac{\zeta}{\varphi}\right) + \frac{1}{\varphi} f'_0\left(\frac{\zeta}{\varphi}\right) \frac{\zeta}{\varphi^2}}{\frac{1}{\varphi} f_0\left(\frac{\zeta}{\varphi}\right)} \\ &= \frac{-1}{\varphi} \left(1 + \frac{f'_0\left(\frac{\zeta}{\varphi}\right) \frac{\zeta}{\varphi}}{f_0\left(\frac{\zeta}{\varphi}\right)}\right) \\ &= \frac{-1}{\varphi} \left(1 + \frac{f'_0(z) z}{f_0(z)}\right) \\ &= \frac{1}{\varphi} s_2(\zeta). \end{aligned}$$

Therefore,

$$\frac{\partial L}{\partial \varphi} = \sum_{t=1}^T \frac{1}{\varphi} s_2(\zeta_t)$$

and

$$\begin{aligned} E \left[\frac{\partial L}{\partial \varphi} \frac{\partial L}{\partial \theta} \right] &= \frac{1}{\varphi} I_2(f) E \left[\frac{\partial \log \lambda_t}{\partial \theta} \right] \\ E \left[\frac{\partial L}{\partial \varphi} \frac{\partial L}{\partial \delta} \right] &= \frac{1}{\varphi} I_2(f) E \left[\frac{1}{\lambda_t} \right]. \end{aligned}$$

In this case, the efficient score function for φ is

$$L_\varphi^* = \sum_{t=1}^T \frac{1}{\varphi} s_2(\zeta_t) \left(1 - \frac{E\left(\frac{1}{\lambda_t}\right)}{E\left(\frac{1}{\lambda_t^2}\right)} \frac{1}{\lambda_t} \right).$$

C.3 Unknown f

We next consider the semiparametric case where f is of unknown form but has unit mean.

In this case for any parameterization φ

$$\begin{aligned}\int f_\varphi(\zeta)d\zeta = 1 &\implies \int \frac{\partial f_\varphi(\zeta)/\partial\varphi}{f_\varphi(\zeta)} f_\varphi(\zeta)d\zeta = 0 \\ \int \zeta f_\varphi(\zeta)d\zeta = 1 &\implies \int \zeta \frac{\partial f_\varphi(\zeta)/\partial\varphi}{f_\varphi(\zeta)} f_\varphi(\zeta)d\zeta = 0.\end{aligned}$$

According to Drost and Werker, the tangent space for f consists of functions τ that satisfy

$$\mathbb{T}_f = \left\{ \tau(\cdot) : \int \zeta^j \tau(\zeta) f(\zeta) d\zeta = 0, \quad j = 0, 1. \right\}$$

If we project

$$L_\theta^* = \sum_{t=1}^T s_2(\zeta_t) \frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right)$$

on to the tangent space for f we have

$$L_\theta^{**} = \sum_{t=1}^T \left(s_2(\zeta_t) + \frac{1 - \zeta_t}{\sigma_\zeta^2} \right) E \left[\frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) \right], \quad (40)$$

because

$$s_2(\zeta) + \frac{1 - \zeta}{\sigma_\zeta^2} \in \mathbb{T}_f$$

since

$$E \left(s_2(\zeta) + \frac{1 - \zeta}{\sigma_\zeta^2} \right) = 0, \quad E \left(\zeta s_2(\zeta) + \frac{\zeta - \zeta^2}{\sigma_\zeta^2} \right) = 0.$$

Since

$$\begin{aligned}- \int \zeta s_2(\zeta) f(\zeta) d\zeta &= \int \zeta f(\zeta) d\zeta + \int \zeta^2 \frac{f'(\zeta)}{f(\zeta)} f(\zeta) d\zeta \\ &= 1 + \int \zeta^2 f'(\zeta) d\zeta \\ &= 1 - 2 \int \zeta f(\zeta) d\zeta \\ &= -1\end{aligned}$$

by integration by parts. Furthermore, $E(\zeta^2) = 1 + \sigma_\zeta^2$ and the result follows.

Therefore, the efficient score function in the model where both g, f are unknown is

$$\begin{aligned}
L_\theta^* &= \sum_{t=1}^T s_2(\zeta_t) \left(\frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) - E \left[\frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) \right] \right) \\
&\quad + \sum_{t=1}^T \frac{\zeta_t - 1}{\sigma_\zeta^2} E \left[\frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) \right] \\
&= \sum_{t=1}^T s_2(\zeta_t) \left(\left(\frac{\partial \log \lambda_t}{\partial \theta} - E \left(\frac{\partial \log \lambda_t}{\partial \theta} \right) - \left(\frac{1}{\lambda_t} - E \left(\frac{1}{\lambda_t} \right) \right) \frac{E \left[\frac{\partial \log \lambda_t}{\partial \theta} \frac{1}{\lambda_t} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) \right) \\
&\quad + \sum_{t=1}^T \frac{\zeta_t - 1}{\sigma_\zeta^2} \left(E \left(\frac{\partial \log \lambda_t}{\partial \theta} \right) - \frac{E \left(\frac{1}{\lambda_t} \right) E \left[\frac{\partial \log \lambda_t}{\partial \theta} \frac{1}{\lambda_t} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right).
\end{aligned}$$

Let $\theta = (\beta, \gamma)$ and $\eta = (\theta^\top, \sigma_\zeta^2)^\top$. The log likelihood function for η based on known f, g is

$$\begin{aligned}
L(\eta | \ell_1, \dots, \ell_T) &= - \sum_{t=1}^T \ln \lambda_t(\theta) - \sum_{t=1}^T \ln g(t/T) - \frac{T}{2} \ln \sigma_\zeta^2 - \sum_{t=1}^T \ln f_0(z_t) \\
z_t &= \frac{\frac{\ell_t}{\lambda_t(\theta)g(t/T)} - 1}{\sigma_\zeta} = \frac{\ell_t - \lambda_t(\theta)g(t/T)}{\sigma_\zeta \lambda_t(\theta)g(t/T)}.
\end{aligned}$$

Let $s_1(z) = -f'_0(z)/f_0(z)$ and $s_2(z) = -((f'_0(z)/f_0(z))z + 1)$, then

$$\frac{\partial L}{\partial \theta} = \sum_{t=1}^T \frac{\partial \ln \lambda_t(\theta)}{\partial \theta} s_2(z_t)$$

$$\frac{\partial L}{\partial \sigma_\zeta^2} = \sum_{t=1}^T s_2(z_t).$$

We have

$$E \left(\frac{\partial L}{\partial \eta} \frac{\partial L}{\partial \eta^\top} \right) = I_2(f)M, \quad M = \begin{pmatrix} E \left(\frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta^\top} \right) & E \left(\frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta} \right) \\ E \left(\frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta^\top} \right) & 1 \end{pmatrix}.$$

The efficient score for θ in the presence of unknown σ_ζ^2 but known g, f_0 is

$$\frac{\partial L^*}{\partial \theta} = \sum_{t=1}^T \left(\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} - E \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \right] \right) s_2(z_t).$$

It follows that the asymptotic variance of the MLE procedure based on known f, g is

$$\sqrt{T}(\hat{\eta} - \eta) \implies N(0, I_2(f)^{-1}M^{-1}).$$

We can estimate $M, I_2(f)$ by

$$\widehat{M} = \begin{pmatrix} \frac{1}{T} \sum_{t=1}^T \frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta^\top} & \frac{1}{T} \sum_{t=1}^T \frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta} \\ \frac{1}{T} \sum_{t=1}^T \frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta^\top} & 1 \end{pmatrix}$$

$$I_2(f) = \frac{1}{T} \sum_{t=1}^T \widehat{s}_2^2(z_t), \quad \widehat{s}_2(z) = - \left(\frac{\widehat{f}'(z)}{\widehat{f}(z)} z + 1 \right).$$

We next show that the score function $\partial L^*/\partial \theta$ is orthogonal to the scores for f, g so that asymptotically, the semiparametric procedure with estimated f, g is first order equivalent to the MLE with known f, g .

We suppose that

$$\ln g(u) = \sum_{j=0}^{\infty} \psi_j H_j(u)$$

$$\ln f_0(v) = \sum_{j=0}^{\infty} \alpha_j \Gamma_j(v).$$

for some basis functions H_j, Γ_j . Then

$$L(\theta, \eta; \psi, \alpha | \ell_1, \dots, \ell_T) = - \sum_{t=1}^T \ln \lambda_t(\theta) - \sum_{t=1}^T \sum_{j=0}^M \psi_j H_j(t/T) - \sum_{t=1}^T \sum_{j=0}^N \alpha_j \Gamma_j(z_t)$$

$$z_t = \frac{\ell_t}{\lambda_t(\theta) \exp \left(\sum_{j=0}^M \psi_j H_j(t/T) \right)}.$$

$$\frac{\partial L}{\partial \theta} = \sum_{t=1}^T \frac{\partial \ln \lambda_t(\theta)}{\partial \theta} s_2(z_t)$$

$$\frac{\partial L}{\partial \psi} = \sum_{t=1}^T H(t/T) s_2(z_t)$$

$$\frac{\partial L}{\partial \alpha} = - \sum_{t=1}^T \Gamma(z_t),$$

where $H = (H_0, \dots, H_M)^\top$ and $\Gamma = (\Gamma_0, \dots, \Gamma_N)^\top$ and $s_2(z_t) = \sum_{j=0}^N \alpha_j \Gamma'_j(z_t) z_t - 1$. We have

$$\begin{aligned} \frac{1}{T} E \left[\frac{\partial L}{\partial \theta} \frac{\partial L}{\partial \theta^\top} \right] &= E \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \frac{\partial \ln \lambda_t(\theta)}{\partial \theta^\top} \right] I_2(f) \\ \frac{1}{T} E \left[\frac{\partial L}{\partial \psi} \frac{\partial L}{\partial \psi^\top} \right] &= \int H(u) H(u)^\top du I_2(f) = I_M \times I_2(f) \\ \frac{1}{T} E \left[\frac{\partial L}{\partial \alpha} \frac{\partial L}{\partial \alpha^\top} \right] &= E \left[\Gamma(z_t) \Gamma(z_t)^\top \right] \\ \frac{1}{T} E \left[\frac{\partial L}{\partial \theta} \frac{\partial L}{\partial \psi^\top} \right] &= E \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \right] \int H(u)^\top du I_2(f) = I_2(f) E \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \right] (1, 0, \dots, 0) \\ \frac{1}{T} E \left[\frac{\partial L}{\partial \theta} \frac{\partial L}{\partial \alpha^\top} \right] &= E \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \right] E \left[s_2(z_t) \Gamma(z_t)^\top \right] \\ \frac{1}{T} E \left[\frac{\partial L}{\partial \psi} \frac{\partial L}{\partial \alpha^\top} \right] &= (1, 0, \dots, 0)^\top E \left[s_2(z_t) \Gamma(z_t)^\top \right]. \end{aligned}$$

The efficient score of θ wrt ψ is

$$s_{\theta|\psi}^* = \sum_{t=1}^T \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} - E \left(\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \right) \right] s_2(z_t)$$

and this score function is orthogonal to $\partial L / \partial \alpha$.

The efficient score of θ wrt α is

$$s_{\theta|\alpha}^* = \sum_{t=1}^T \frac{\partial \ln \lambda_t(\theta)}{\partial \theta} s_2(z_t) - E \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \right] E \left[s_2(z_t) \Gamma(z_t)^\top \right] \left(E \left[\Gamma(z_t) \Gamma(z_t)^\top \right] \right)^{-1} \times \sum_{t=1}^T \Gamma(z_t)$$

and this score function is orthogonal to $\partial L / \partial \alpha$.

It follows that the efficient score for θ in the presence of unknown σ_ζ^2 but known g, f_0 is

$$\frac{\partial L^*}{\partial \theta} = \sum_{t=1}^T \left(\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} - E \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \right] \right) s_2(z_t).$$

D Risk premium regressions

We consider the population regression model

$$R_{mt} - g(t/T) = a + b\lambda_t + c\zeta_t + \varepsilon_t,$$

where in practice we replace $g(\cdot)$ by $\widehat{g}(\cdot)$ and λ_t, ζ_t by $\lambda_t(\widehat{\theta}), \zeta_t(\widehat{\theta})$. This does not affect consistency but does affect the limiting distribution and hence standard errors. The dependent variable effect takes care of itself, the main issue is around the estimated covariate. We argue as follows. Suppose that

$$y_t = \beta^\top x_t(\theta_0) + \varepsilon_t.$$

By a Taylor expansion we have

$$x_t(\widehat{\theta}) \simeq x_t(\theta_0) + \frac{\partial x_t(\theta_0)}{\partial \theta} (\widehat{\theta} - \theta_0).$$

We also have an expansion for our estimators of the form

$$\widehat{\theta} - \theta_0 = \frac{1}{T} \sum_{t=1}^T \psi_t(\theta_0) + e_t,$$

where $\psi_t(\theta_0)$ are mean zero and the sum satisfies a CLT, while e_t is of smaller order. Let

$$x_t^* = x_t(\widehat{\theta}) - \frac{\partial x_t(\widehat{\theta})}{\partial \theta} \frac{1}{T} \sum_{t=1}^T \psi_t(\widehat{\theta}).$$

Then we regress $\widehat{y}_t$ on x_t^* and use the linear regression standard errors.

E Other tables and figures

E.1 Estimation based on conditional moment restrictions

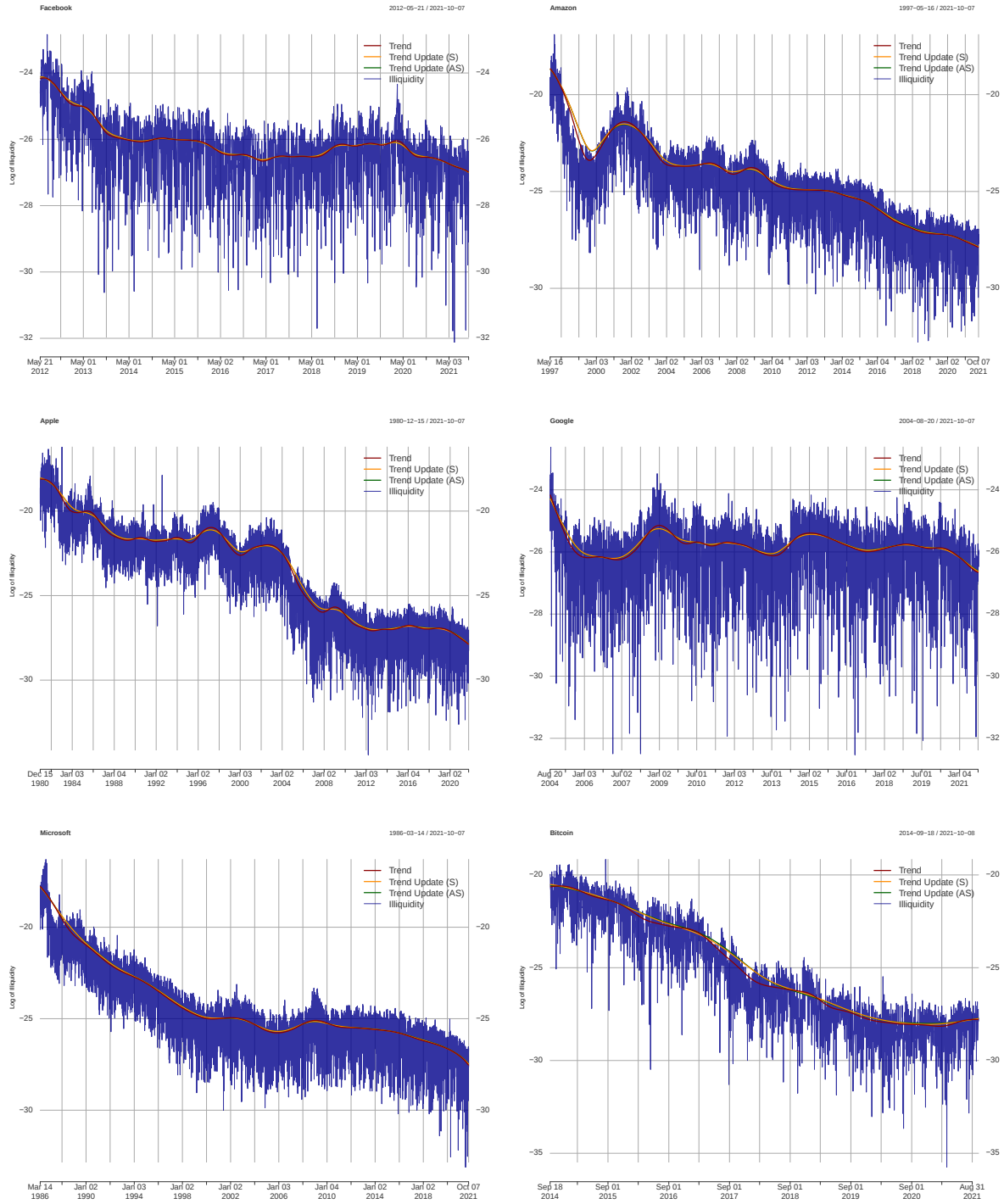


Figure 7: Fab 5 and Bitcoin log illiquidity and updated trend function based on the GMM estimator of λ_t parameters. The red curve corresponds to the initial estimate of the trend function and the yellow and green curves correspond to the updated estimates in, respectively, the symmetric and asymmetric specifications of λ_t .

E.2 Estimation: i.i.d. error term with parametric density

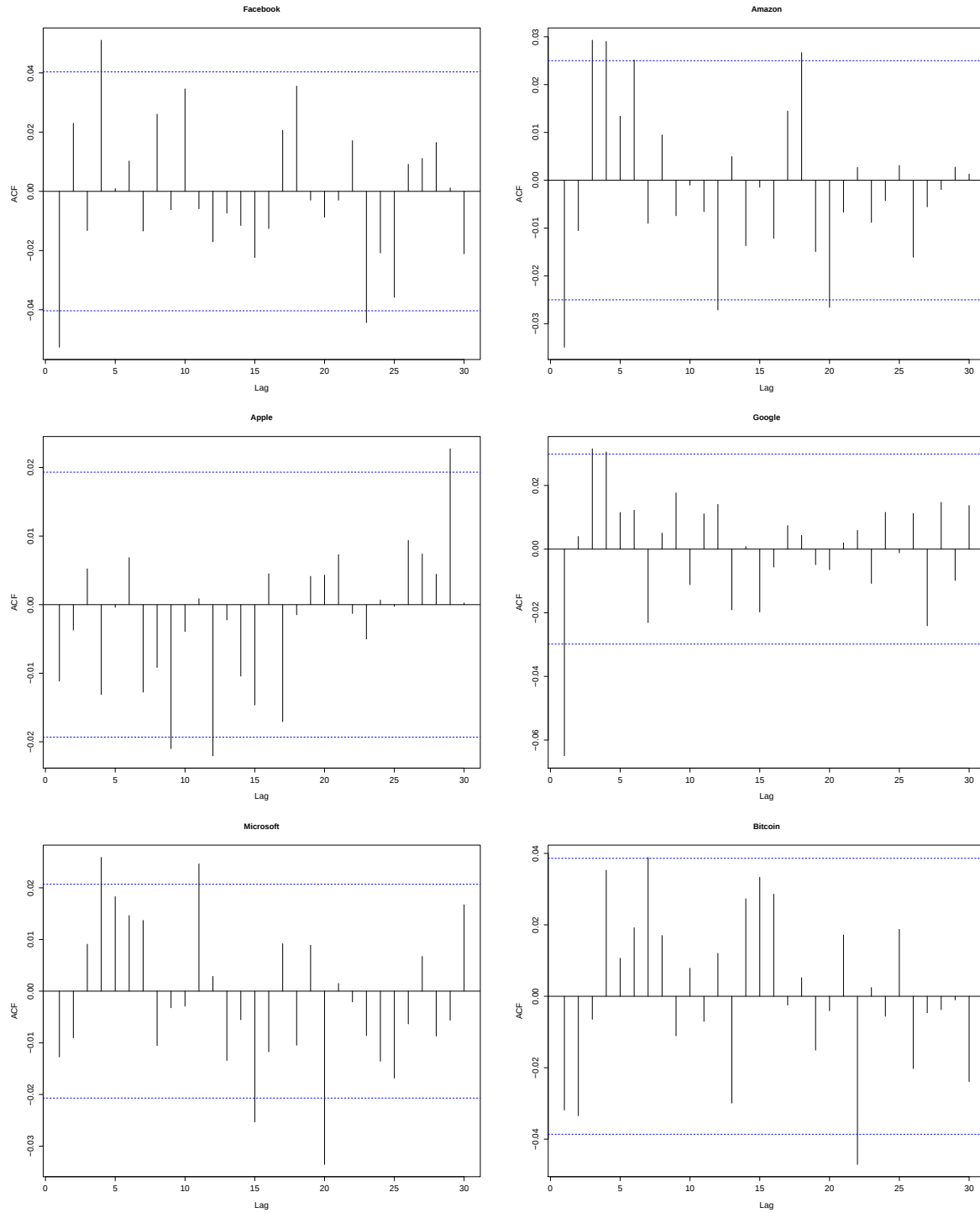


Figure 8: ACF of ζ_t under the symmetric specification for λ_t .

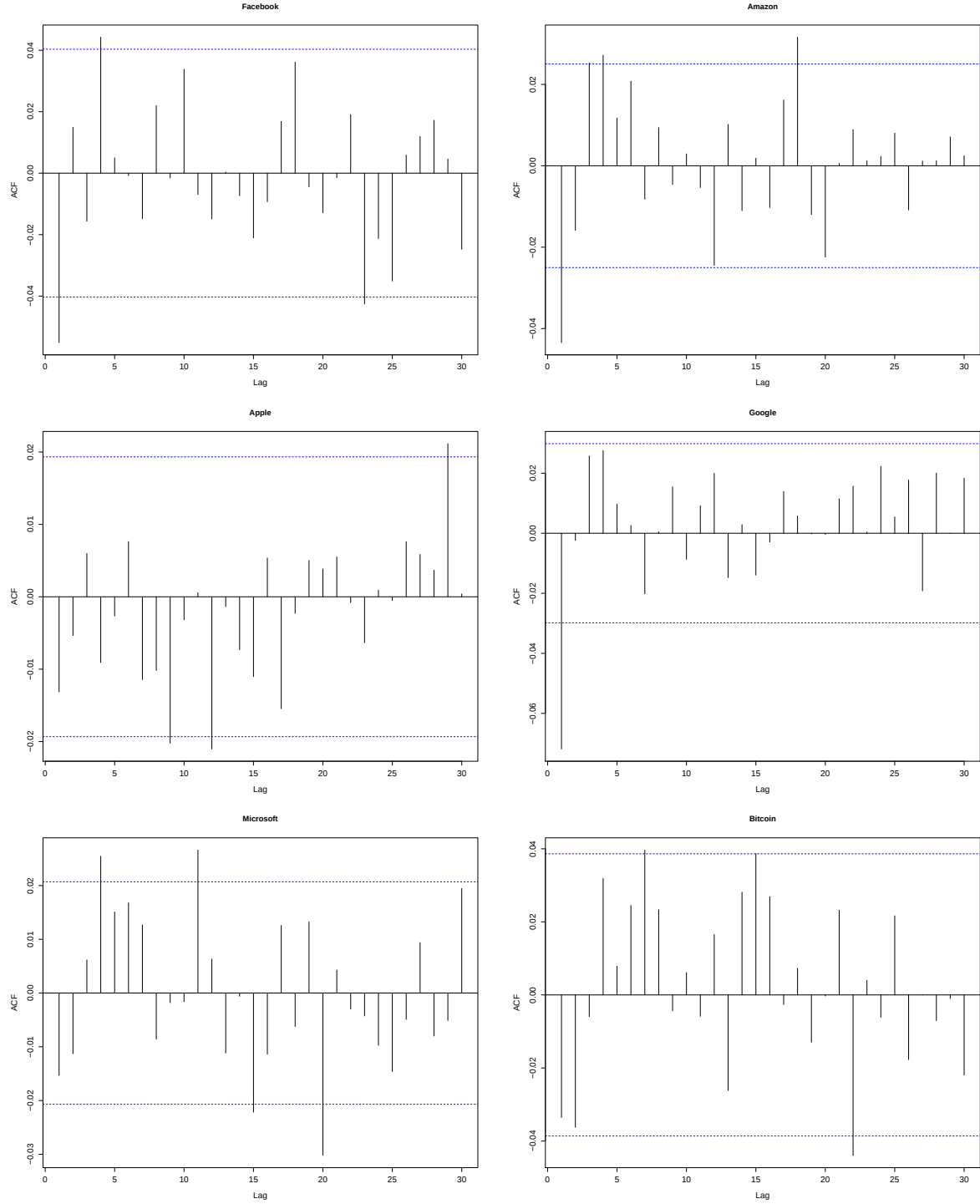


Figure 9: ACF of ζ_t under the asymmetric specification for λ_t .

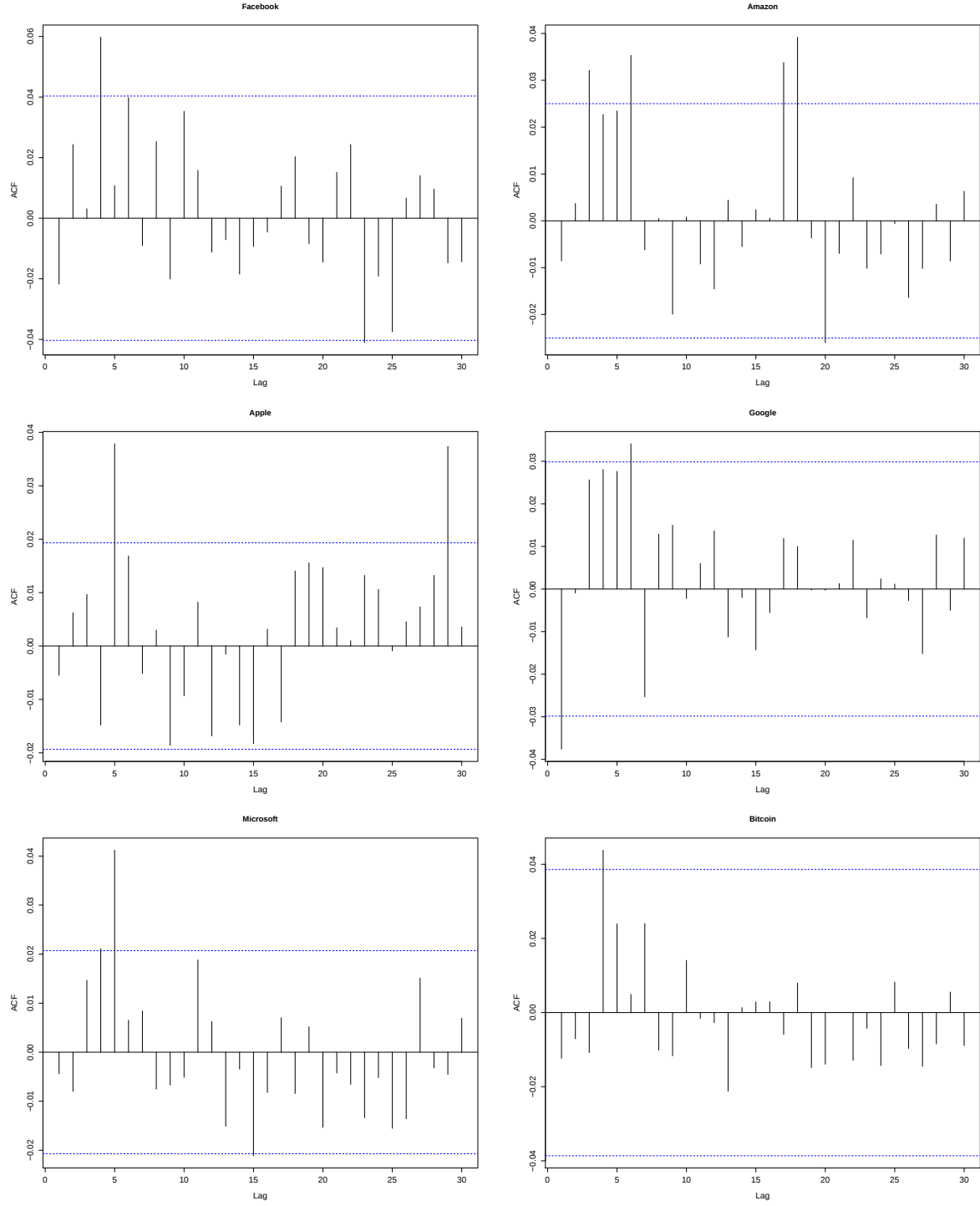


Figure 10: ACF of ζ_t^2 under the symmetric specification for λ_t .

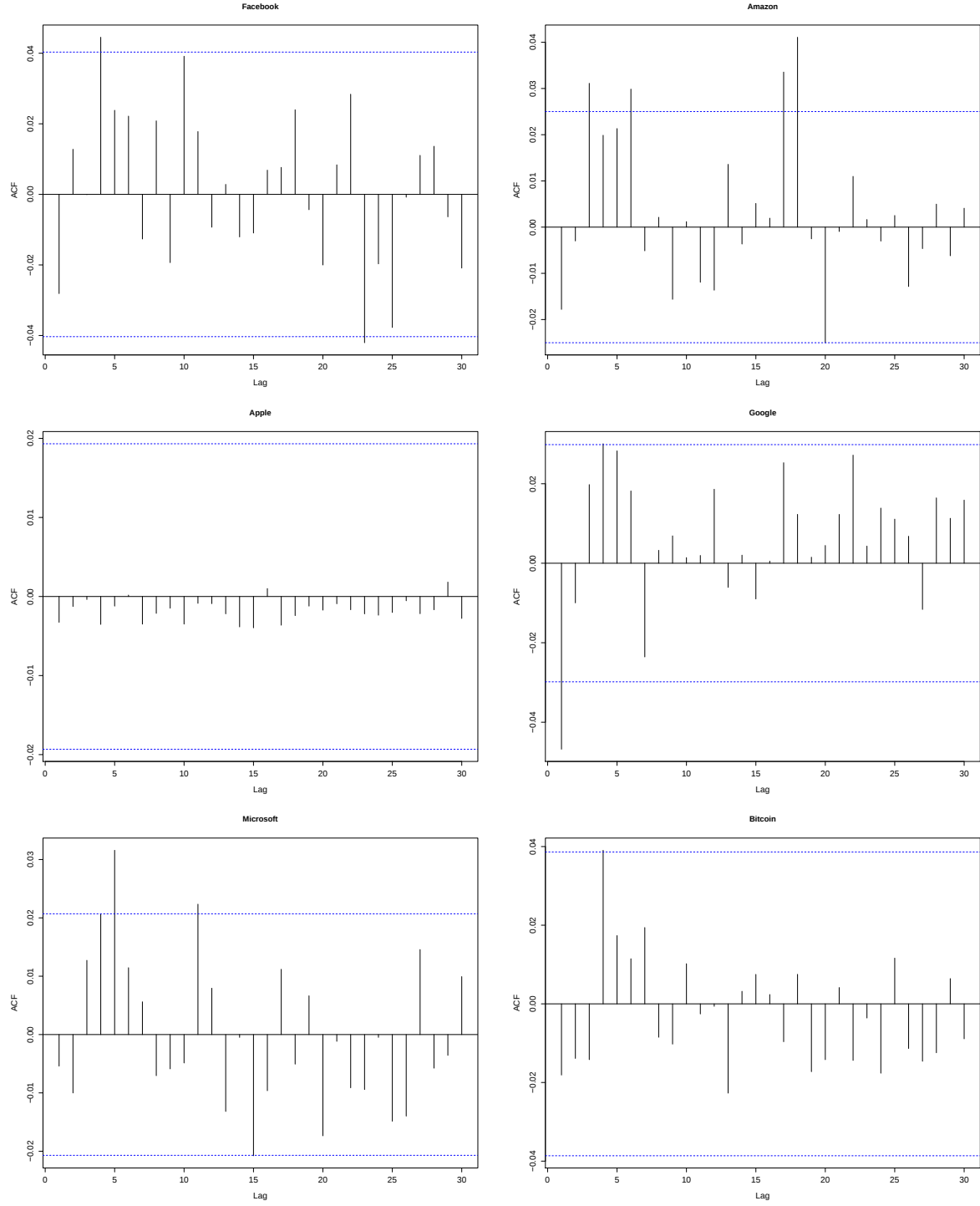


Figure 11: ACF of ζ_t^2 under the asymmetric specification for λ_t .

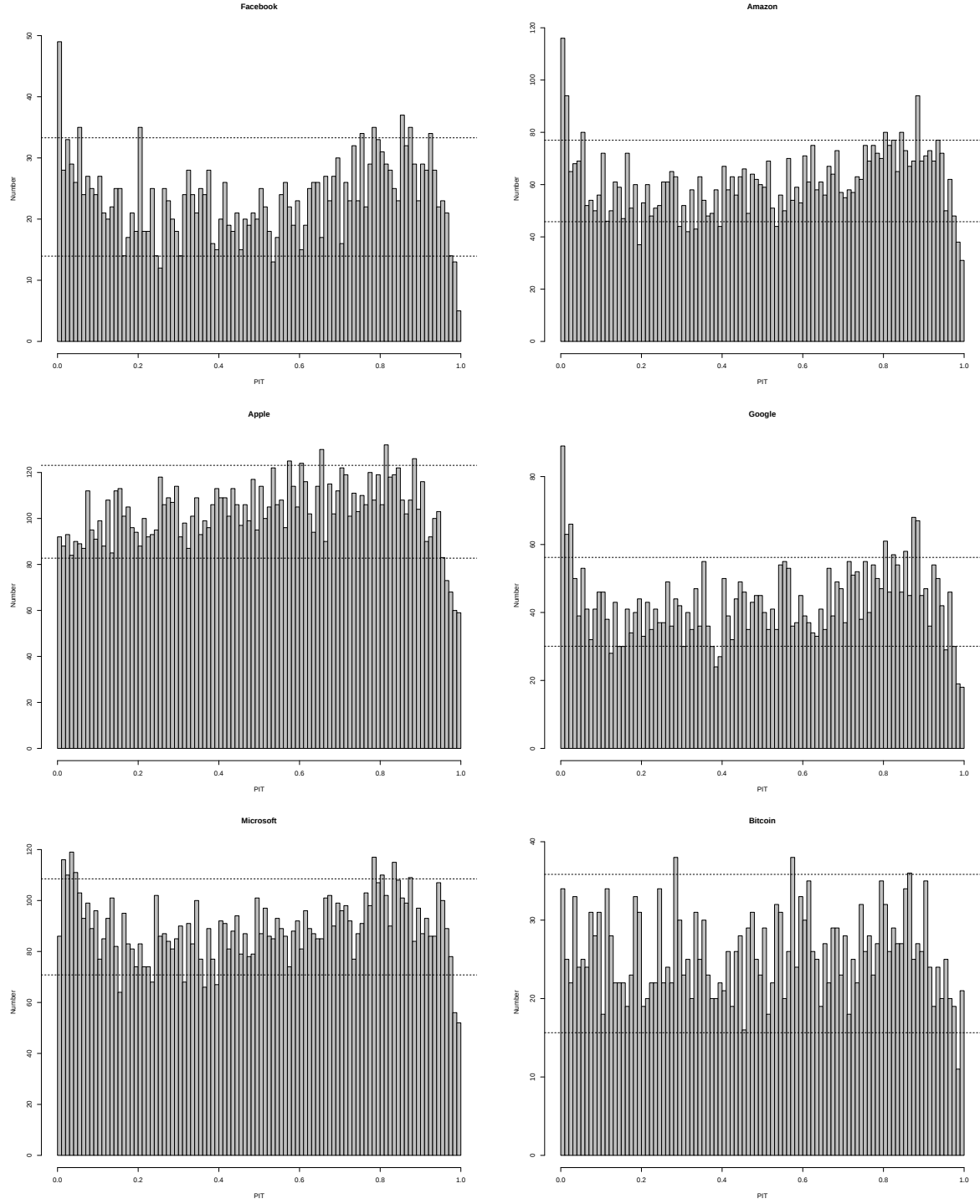


Figure 12: Probability integral transform (PIT) of ζ_t under the symmetric specification for λ_t .

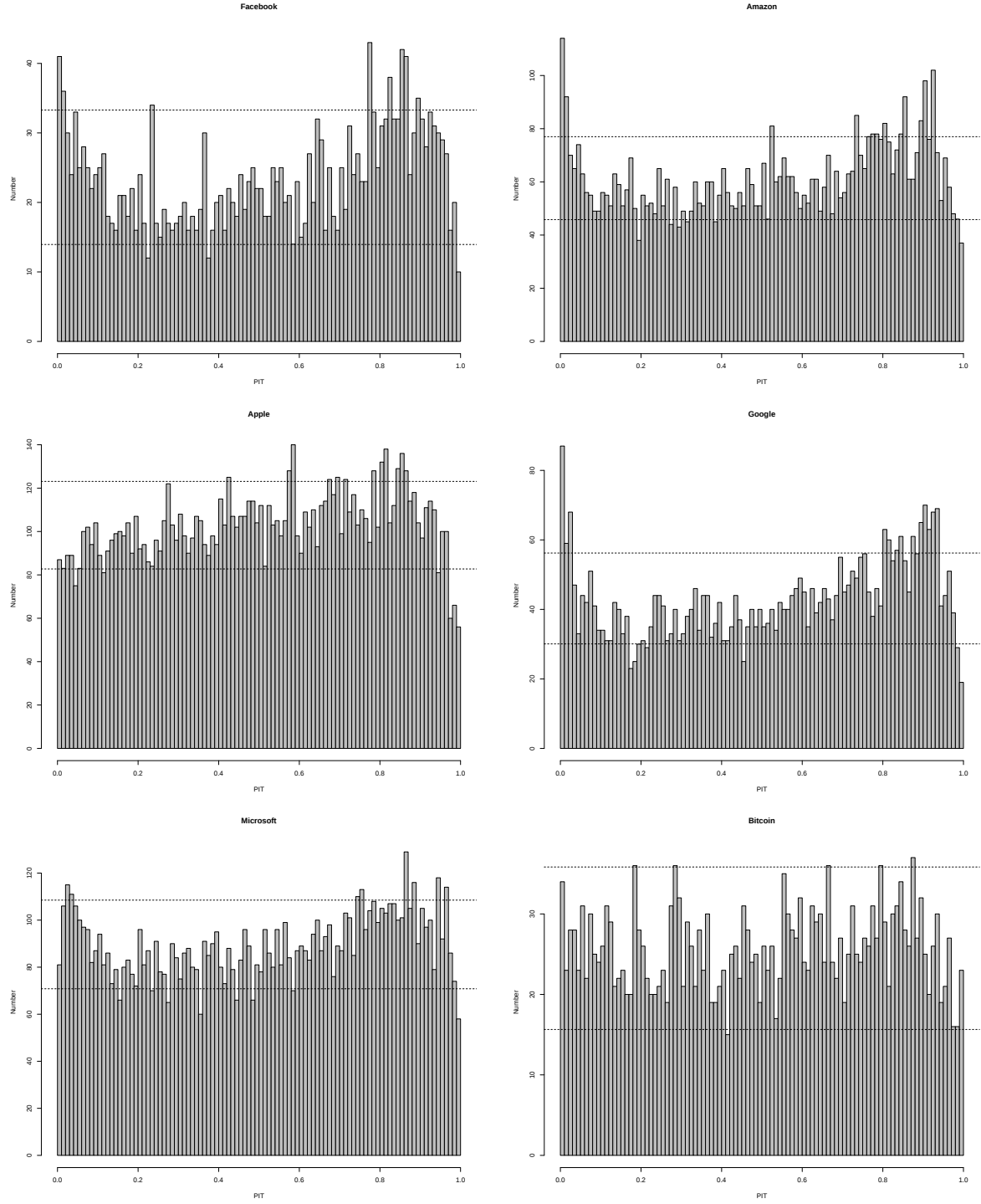


Figure 13: Probability integral transform (PIT) of ζ_t under the asymmetric specification for λ_t .

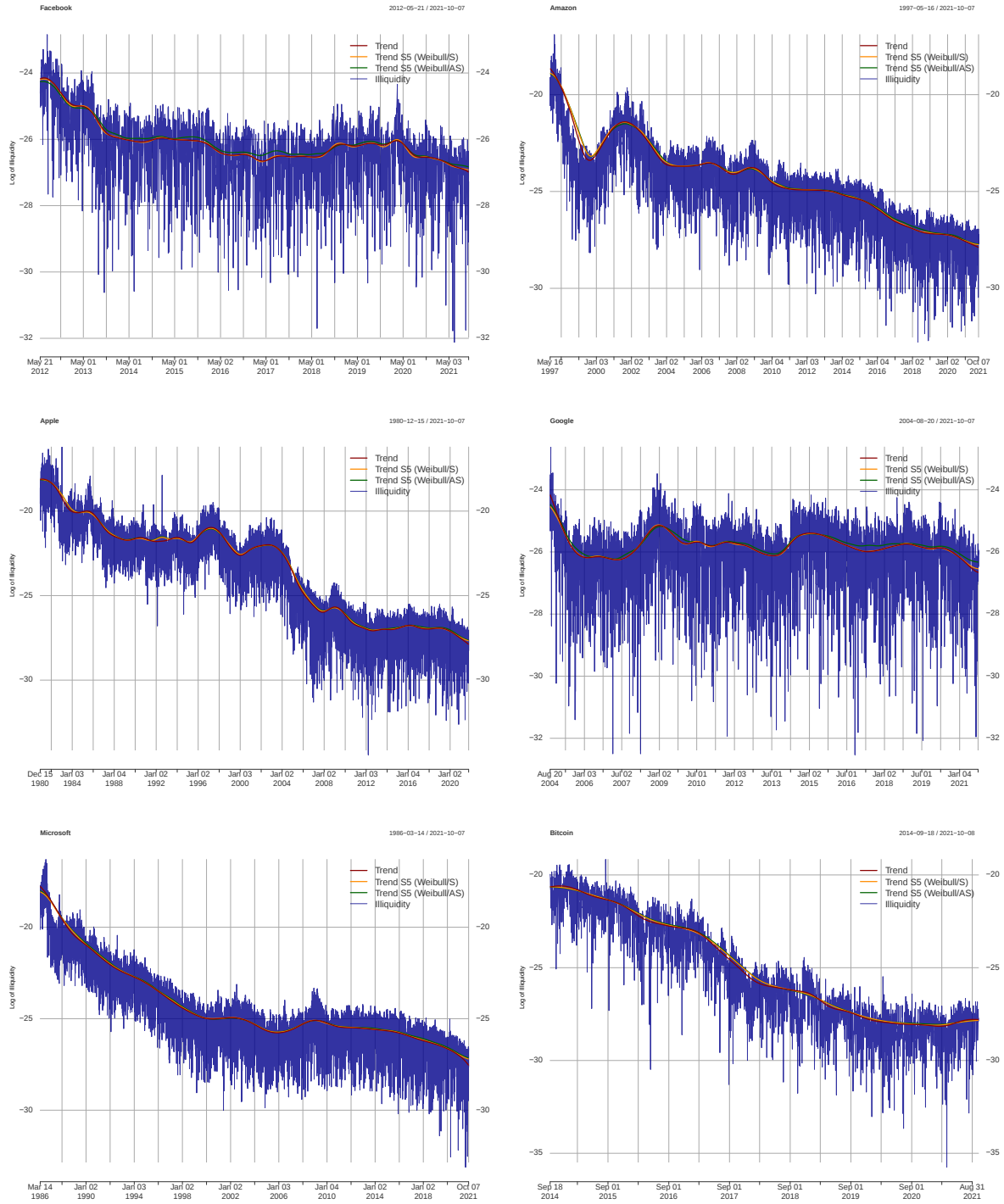


Figure 14: Fab 5 and Bitcoin log illiquidity and updated trend function based on the semiparametric ML estimator of λ_t parameters where the error term ζ_t follows a Weibull distribution. The red curve corresponds to the initial estimate of the trend function and the yellow and green curves correspond to the updated estimates in, respectively, the symmetric and asymmetric specifications of λ_t .

E.3 Estimation: i.i.d. error term with nonparametric density

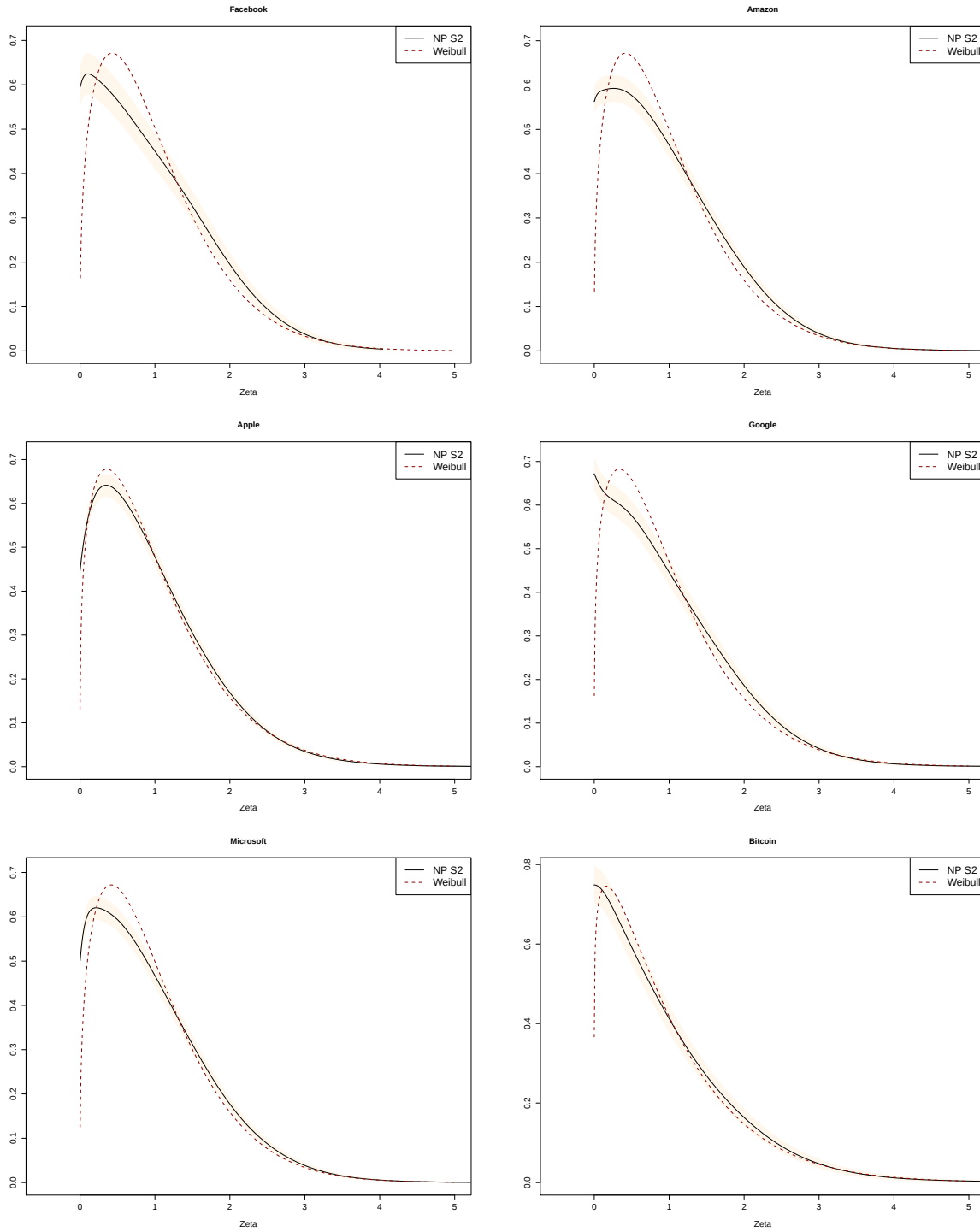


Figure 15: Comparison between the kernel density estimate of ζ_t (solid line) and the Weibull density (dashed line) under the symmetric specification for λ_t .

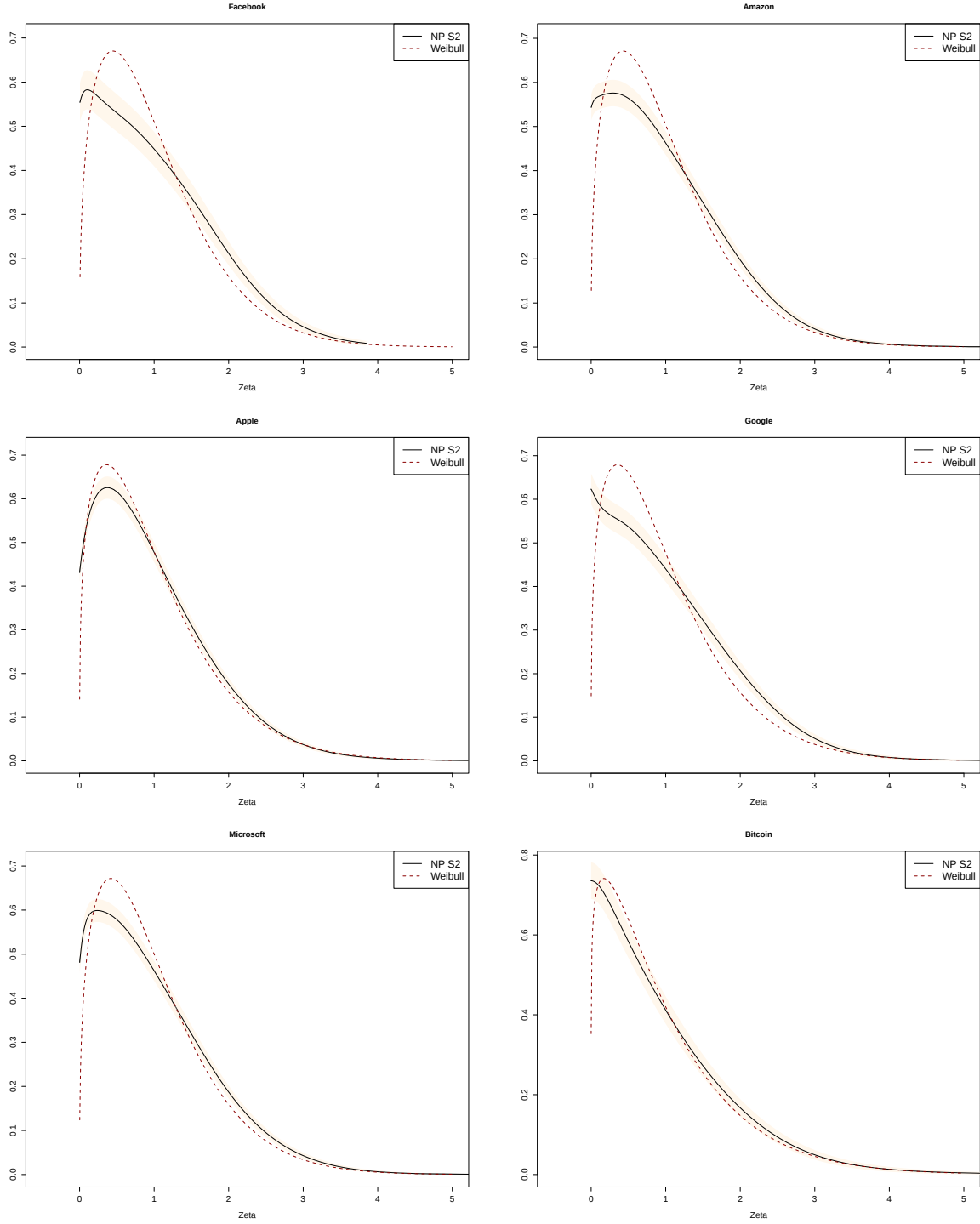


Figure 16: Comparison between the kernel density estimate of ζ_t (solid line) and the Weibull density (dashed line) under the asymmetric specification for λ_t .

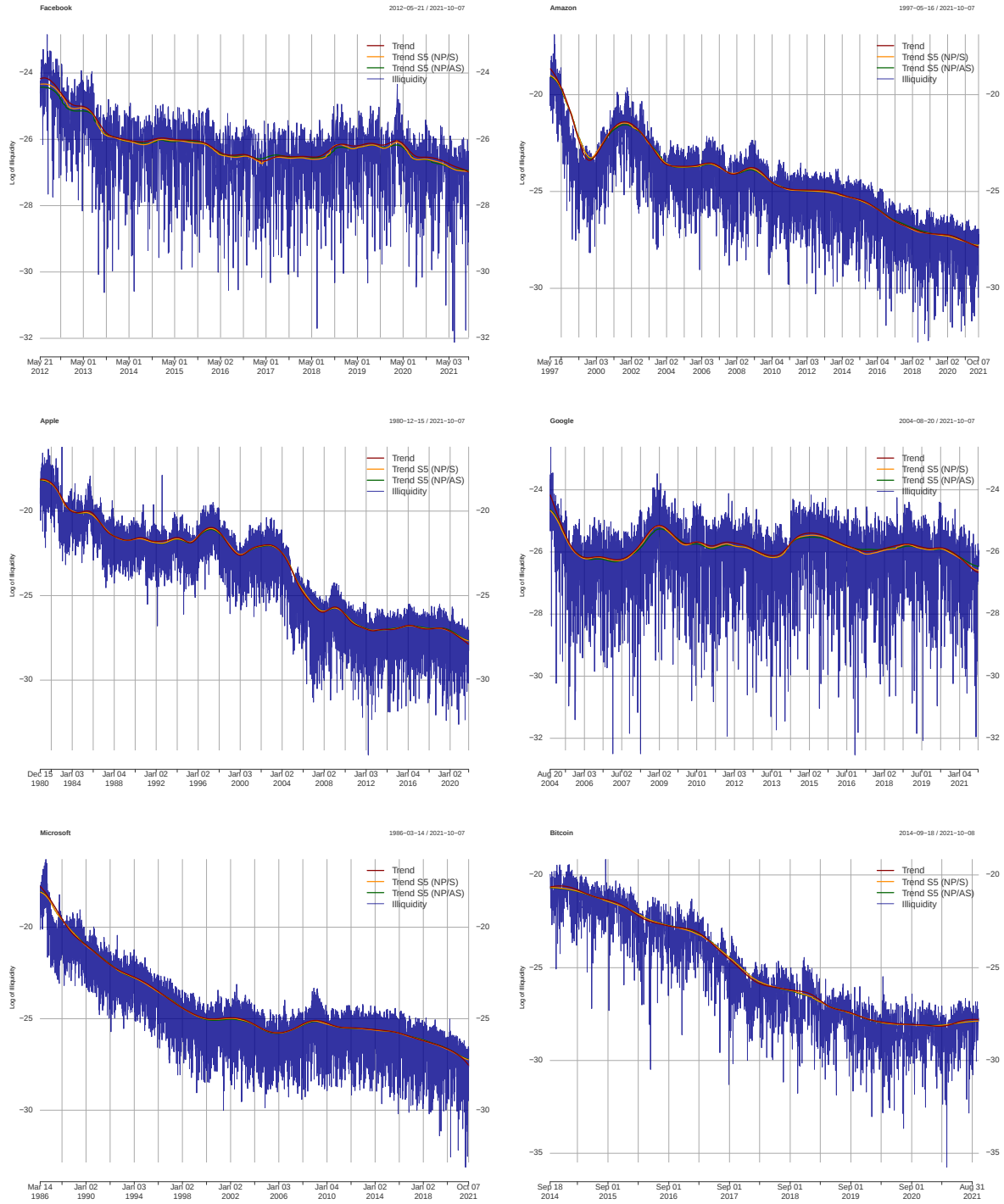


Figure 17: Fab 5 and Bitcoin log illiquidity and updated trend function based on the semiparametric ML estimator of λ_t parameters where the density of the error term ζ_t is estimated nonparametrically. The red curve corresponds to the initial estimate of the trend function and the yellow and green curves correspond to the updated estimates in, respectively, the symmetric and asymmetric⁷³ specifications of λ_t .

E.4 Risk premium

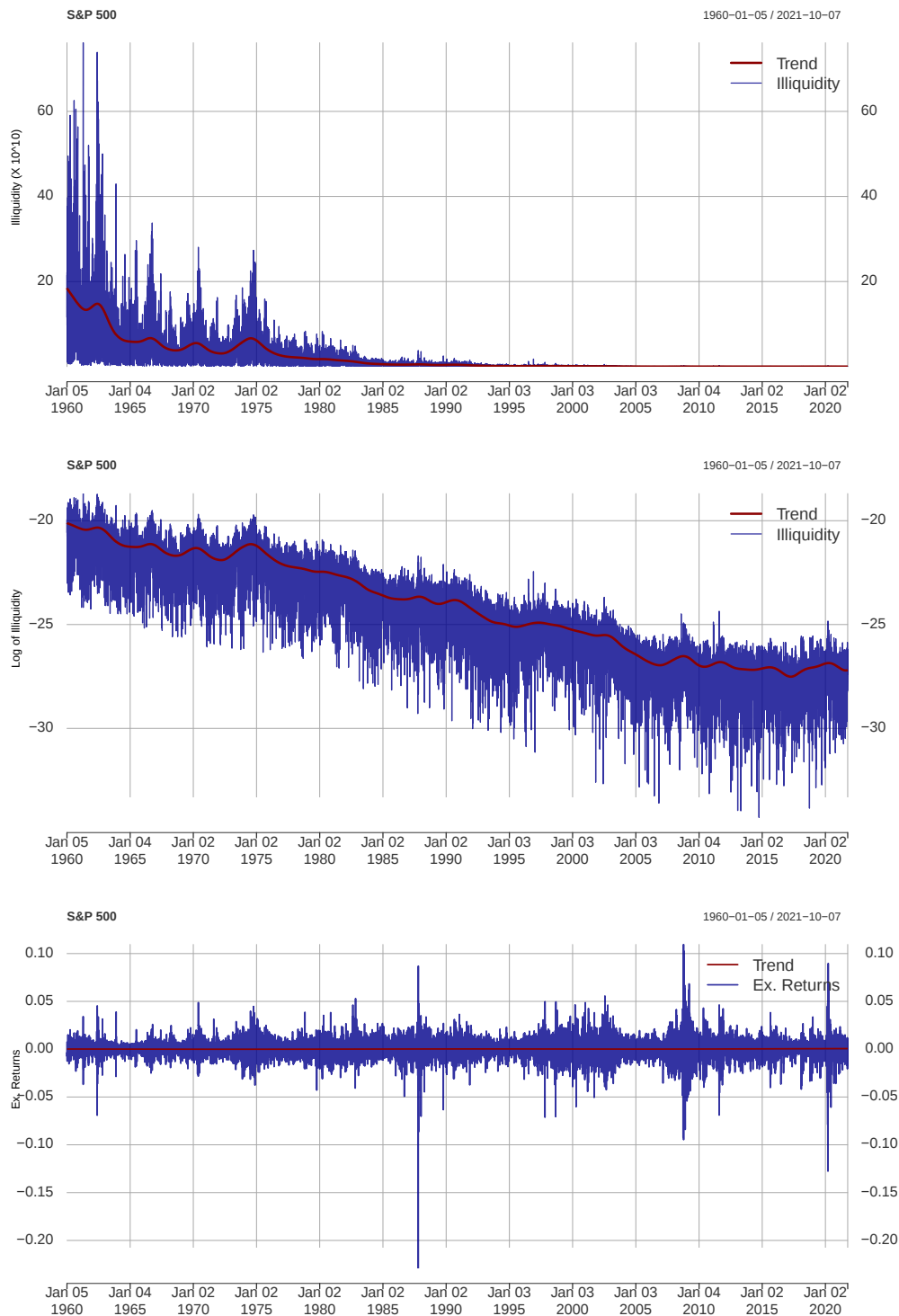


Figure 18: Daily (log) illiquidity series and return data.

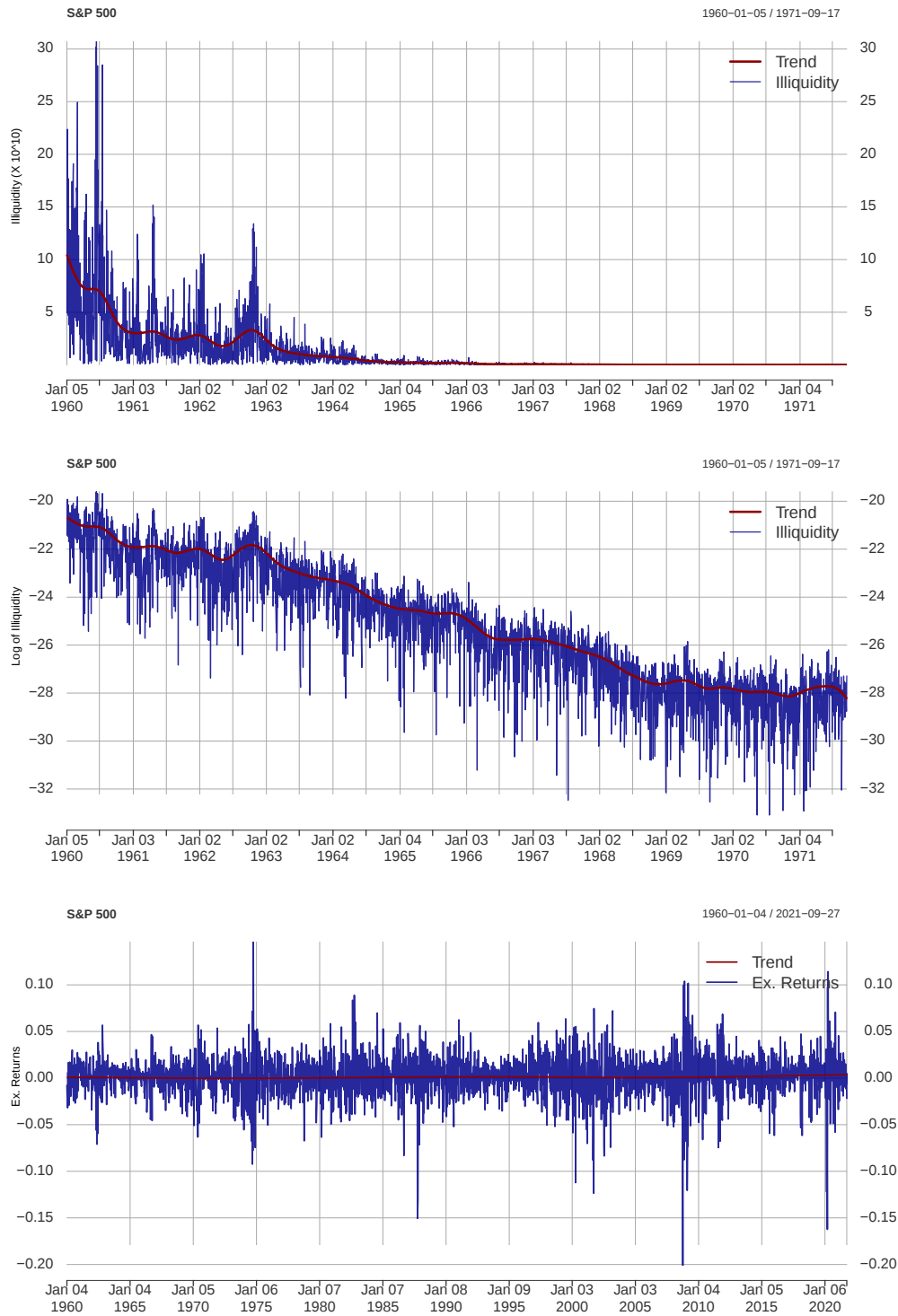


Figure 19: Weekly (log) illiquidity series and return data.

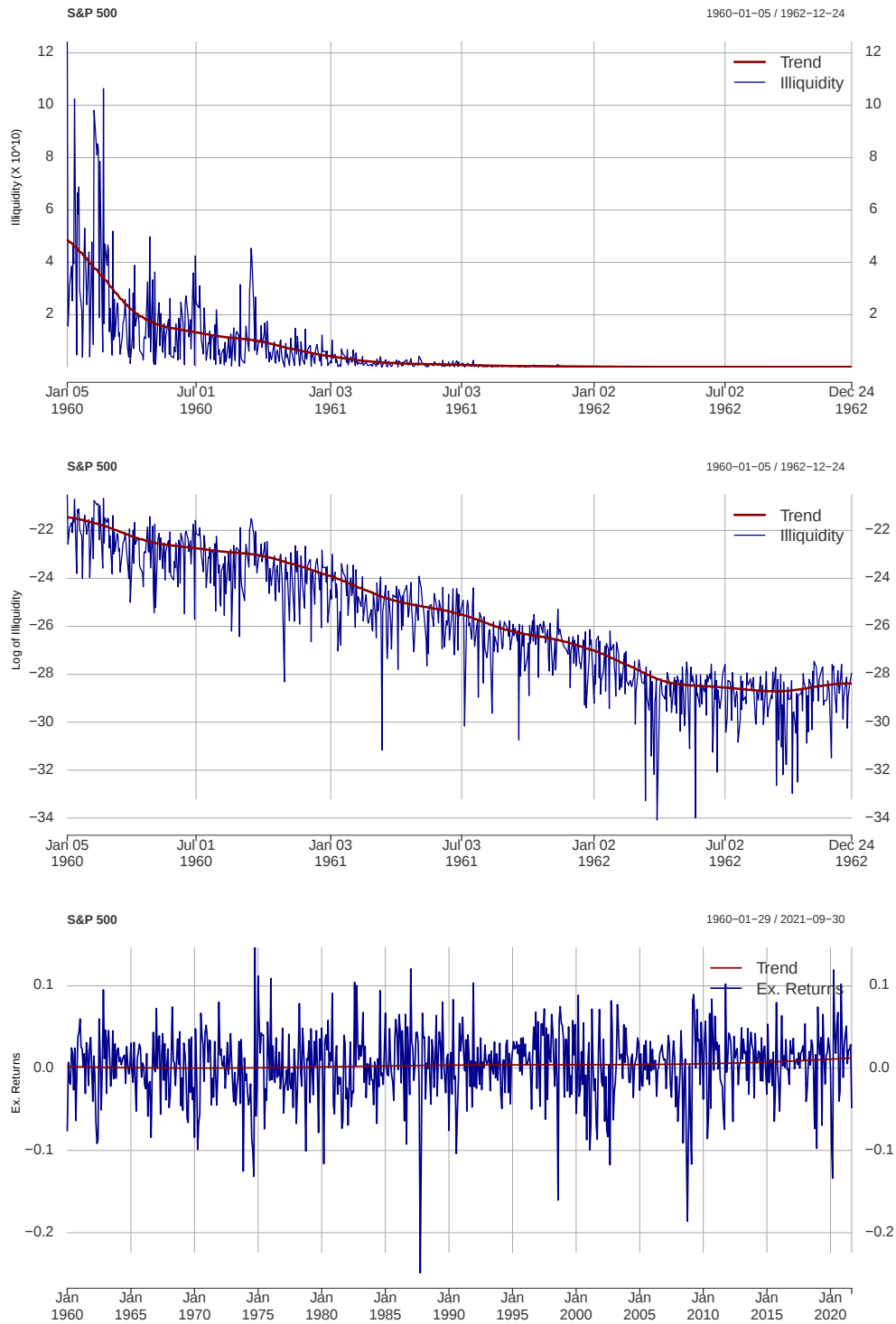


Figure 20: Monthly (log) illiquidity series and return data.

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.
- BEDDINGTON, J., FURSE, C., BOND, P., CLIFF, D., GOODHART, C., HOUSTOUN, K., LINTON, O., ZIGRAND, J.-P., et al. (2012): *Foresight: The future of computer trading in financial markets: Final project report*. Tech. rep. London School of Economics and Political Science, LSE Library.
- BROGAARD, J. (2010): “High frequency trading and its impact on market quality”.
- CHEN, X. (2007): “Large sample sieve estimation of semi-nonparametric models”. *Handbook of econometrics* 6, pp. 5549–5632.
- CHEN, X., LINTON, O. AND VAN KEILEGOM, I. (2003): “Estimation of semiparametric models when the criterion function is not smooth”. *Econometrica* 71.5, pp. 1591–1608.
- COPELAND, T. E. (1979): “Liquidity changes following stock splits”. *The Journal of Finance* 34.1, pp. 115–141.
- DELGADO, M. A. AND HIDALGO, J. (2000): “Nonparametric inference on structural breaks”. *Journal of Econometrics* 96.1, pp. 113–144.
- DROST, F. C. AND KLAASSEN, C. A. (1997): “Efficient estimation in semiparametric GARCH models”. *Journal of Econometrics* 81.1, pp. 193–221.
- 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”. *Review of Economics and Statistics* 94.1, pp. 222–223.
- FAN, J. AND CHEN, J. (1999): “One-step local quasi-likelihood estimation”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.4, pp. 927–943.

- 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.
- HAFNER, C. M. AND LINTON, O. (2010): “Efficient estimation of a multivariate multiplicative volatility model”. *Journal of econometrics* 159.1, pp. 55–73.
- HASBROUCK, J. (2009): “Trading Costs and Returns for U.S. 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.
- LINTON, O. AND XIAO, Z. (2001): “Second-order approximation for adaptive regression estimators”. *Econometric Theory* 17.5, pp. 984–1024.
- LU, Z. AND LINTON, O. (2007): “Local linear fitting under near epoch dependence”. *Econometric Theory* 23.1, pp. 37–70.
- O’HARA, M. AND YE, M. (2011): “Is market fragmentation harming market quality?” *Journal of Financial Economics* 100.3, pp. 459–474.
- VOGT, M. (2012): “Nonparametric regression for locally stationary time series”. *The Annals of Statistics* 40.5, pp. 2601–2633.
- YATCHEW, A. (2000): “Scale economies in electricity distribution: A semiparametric analysis”. *Journal of applied Econometrics* 15.2, pp. 187–210.