UNIVERSITY OF CAMBRIDGE
Faculty of Economics

JANEWAY INSTITUTE CAMBRIDGE

CAMBRIDGE WORKING PAPERS IN ECONOMICS
JANEWAY INSTITUTE WORKING PAPERS

Dynamic Autoregressive Liquidity (DArLiQ)

Christian M. Hafner
UCLouvain

Oliver B. Linton
University of Cambridge

Linqi Wang
UCLouvain

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 our 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.

Reference Details

2214	Cambridge Working Papers in Economics
2022/06	Janeway Institute Working Paper Series
Published	23 February 2022
Revised	16 September 2022
Key Words	Kernel, Nonparametric Estimation, Semiparametric Model, Stock Splits, Structural Change
JEL Codes	C12, C14
Websites	www.econ.cam.ac.uk/cwpe www.janeway.econ.cam.ac.uk/working-papers

Dynamic Autoregressive Liquidity (DArLiQ)

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

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

[b]Faculty of Economics, University of Cambridge

September 16, 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 our 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: Kernel; Nonparametric Estimation; Semiparametric Model; Stock Splits; Structural Change.

JEL Classification: C12, C14

*Email: christian.hafner@uclouvain.be.

†Email: obl20@cam.ac.uk.

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

1 Introduction

Liquidity is a fundamental property of a well-functioning market, and lack of liquidity is generally at the heart of many financial crises and disasters. Common ways of measuring liquidity using high-frequency data include bid-ask spreads, effective spreads, realized spreads, depth, weighted depth, and transaction volume. There is a large literature that uses such measures to compare market quality across markets, across time, and before and after interventions of various sorts. For example, it has been a big part of the debate around high frequency trading, i.e., whether such trading activity has improved or degraded market liquidity, see e.g. 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 error model (MEM) that contains a nonparametric long-run trend and a parametric short-run autoregressive process as in Engle et al. (2012). The trend part is important for many datasets where liquidity has improved in a secular fashion such as the S&P 500 over the last hundred years and Bitcoin over the much more recent period of its operation. The nonparametric trend is comparable with the conventional monthly averaged measure 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, which is consistent under minimal conditions, as well as through a semiparametric likelihood procedure that assumes i.i.d. shocks. In the latter approach we consider two cases, one where the shock distribution is parametric such as the Burr distribution and a further case in which the shock distribution is not specified and is treated nonparametrically. In all cases, the nonparametric trend affects the limiting distribution of the estimators

of the finite dimensional parameters in a non trivial way, and correct standard errors need to account for that. We develop the distribution theory for the three cases to enable valid inference. In the cases with i.i.d. shocks we establish the semiparametric efficiency bounds and show that our estimators achieve those bounds under correct specification.

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 change versus the alternative 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 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 Burr distribution for the error term, captures some of 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, that is to say there appear to be permanent jumps in the levels of liquidity that are not accompanied by additional dynamic adjustment.

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).

Related Literature. The Multiplicative error model has been applied to many different financial time series including volatility, duration between trades, and transaction volume, see e.g. Engle (2002). The MEM model and its applications and developments over the last 20 years is reviewed in Cipollini and Gallo (2022). The VLAB applies this model and provides regular updates on their website (`https://vlab.stern.nyu.edu/liquidity`) for a number of series according to their specific implementation. We next compare our model with theirs. They fit a parametric model with multiplicative components, the same dynamic model as ours; they also use a "quadratic spline" to capture trends, that is, they include a quadratic function of time. They focus on the iid error case with a chi-squared shock distribution. We treat the trend as a nonparametric function of rescaled time and use local weighting estimators to estimate the trend, as in Hafner and Linton (2010) in which case our trend estimators nest the Amihud low frequency estimators as a special case, whereas the VLAB quadratic spline estimator does not have such a connection. Second, we consider the case where the shock is not iid and use a GMM estimation procedure like Cipollini et al. (2013) except we use only the first moment of the detrended series; this estimation method is robust to higher moment existence and to time variation in higher moments of illiquidity. We also consider the iid shock case but we find two issues. First, the presence of zeros even in the S&P500 series and some individual stocks. We allow for the shock to have a discrete component that can be estimated by the zero frequency separately from the estimation of the continuous part. The second issue is that the shock distribution appears to have heavy tails as quantified by the log rank estimator (tail thickness in the range of four to eight) and so the chi-squared distribution and the Weibull distributions (that are usually used in MEM applications) would appear not to be good choices for the continuous part. Therefore, we consider the Burr distribution that nests the Weibull but allows for Pareto like tails. We also consider a nonparametric shock density for

the continuous part, which is consistent with heavy tails. We also develop the statistical theory necessary to implement inference in our more general class of models and present semiparametric efficiency bounds for the dynamic parameters in the presence of the two nonparametric nuisance functions the trend and shock density function. In that regard, our work extends Drost and Werker (2004) who consider efficiency bounds in the autoregressive conditional duration (ACD) model of Engle and Russell (1998), but without trends. We also develop additional methodology to detect permanent and temporary shifts in illiquidity. We work with kernel smoothing methods throughout. An alternative estimation approach is based on the sieve method, Chen (2007). The advantage of the sieve method is that it only requires a single optimization, albeit one with many parameters to choose.

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. The appendices are delegated to a separate file which is available online as Hafner et al. (2022). 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

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

where R_{t_j} is the stock return and V_{t_j} is the (dollar) trading volume at time t_{t_j}. 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. Barardehi et al. (2021) proposed to replace the close to close return by the overnight component of that return.

Fong et al. (2018) proposed a more general class of liquidity measures based on ratios of functions of volatility to functions of trading volume. Both of these modifications can easily be accommodated in our framework, but we focus on the original Amihud measure as this is currently the most popular approach.

Empirical evidence points to the existence of factors driving low-frequency variations in illiquidity dynamics in addition to higher-frequency variations. To illustrate this point, we plot in Figure 1 the evolution of the daily log Amihud illiquidity measure for the S&P 500 stock market index over the period of 1950–2021. We observe that the illiquidity series exhibit a strong downward trend over time, at least up to 2005. Trends in illiquidity series are not limited to the S&P 500 index used in our illustration. To emphasize how prevalent this feature is across financial markets and to gain more insights into the conditional dynamics of 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 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 in Appendix E.1 of Hafner et al. (2022). The results show that the coefficient estimates for the trend function are significant. One exception is the quadratic term for Microsoft, meaning that this stock exhibits a linear trend over the sample period. Consistent with visual inspection of Figure 1 in Appendix E.1 of Hafner et al. (2022), all estimated polynomial trend functions are overall downward trending. In addition, most of the autoregressive coefficients are statistically significant, indicating some degree of persistence in the stock illiquidity dynamics. Taken together, these evidence motivate our modelling approach for the Amihud illiquidity measure which weakens the requirement on stationarity and we 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 component relevant for the modelling of illiquidity series.

3 The model

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

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

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

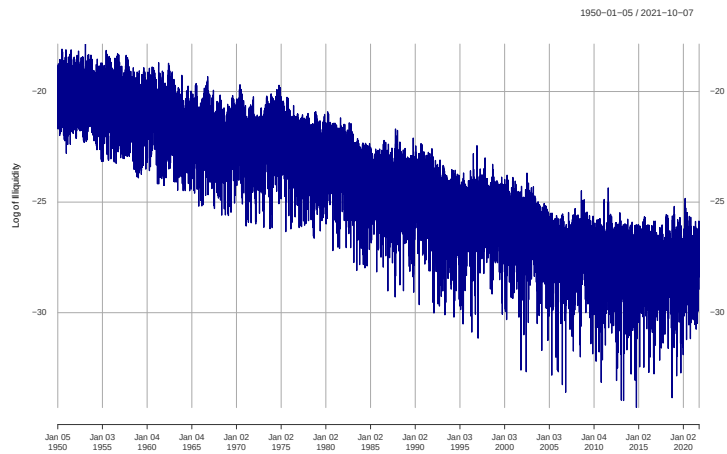


Figure 1: S&P 500 index daily log illiquidity – $\log \ell_t$.

where $g(.)$ is a positive and smooth but unknown function of rescaled time, $\ell_t^* = \ell_t/g(t/T)$ is the rescaled liquidity, and ζ_t is a sequence of non-negative random variables with conditional mean one and finite unconditional variance denoted σ_ζ^2. We present evidence later that the assumption of finite unconditional variance for the shock process is reasonable. Note that $\omega > 0,\ \beta_j \geq 0$ for $j = 1, \ldots, p,\ \gamma_k \geq 0$ for $k = 1, \ldots, q$ are sufficient conditions for $\lambda_t > 0$ with probability one; see Nelson and Cao (1992) for necessary conditions. Furthermore, provided $\sum_{j=1}^p \beta_j + \sum_{k=1}^q \gamma_k < 1$, the process ℓ_t is stationary in mean (and perforce strictly stationary) and follows an ARMA(max(p,q),q) process.

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. We may suppose that the error process ζ_t be i.i.d. with some c.d.f F. Francq and Zakoïan (2006) (Theorems 2 and 3) ensures that the process ℓ_t^* is strictly stationary and geometrically ergodic under our restrictions on β, γ. The i.i.d. assumption can be helpful for estimation but it may also be important for 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 dependency. The process may be initialized from its stationary distribution or from some fixed values $\{\ell_0, \ldots, \ell_{1-q}, \lambda_0^*, \ldots, \lambda_{1-p}^*\}$, which will be discussed below. In the sequel, we suppose that $p = 1, q = 1$ for simplicity, i.e. $\lambda_t = \omega + \beta\lambda_{t-1} + \gamma\ell_{t-1}^*$ and we use an expectation targeting approach to obtain ω by setting $\omega = 1 - \beta - \gamma$. Additionally, we may also consider the specification with asymmetric effect, that is $\lambda_t = \omega + \beta\lambda_{t-1} + \gamma\ell_{t-1}^* + \gamma^- \ell_{t-1}^* \mathrm{I}_{R_{t-1}<0}$ 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 $E\left[\ell^*_{t-1}\mathrm{I}_{R_{t-1}<0} \mid \mathcal{F}_{t-1}\right] = \lambda_t/2$ and we can also use a targeting approach for ω by setting $\omega = 1 - \beta - \gamma - \gamma^-/2$.

We may wish to consider the effects of interventions at some times $t_1, \ldots, t_J$. 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 and zero otherwise. In this case the level of the process λ_t is affected for all $t \geq t_1$, with a flexible effect between t_1 and t_J, but after t_J the effect decreases rapidly as $t - t_J \to \infty$. The null hypothesis of interest here is $\alpha_1 = \cdots = \alpha_J = 0$, in which case the model collapses to Equation (3).

We may allow the possibility of permanent effects (structural change) by allowing the function g to be discontinuous at points $u_1, \ldots, u_M \in (0, 1)$, that is, for a given point u_*

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

are both well defined but $g_-(u_*)$ may not be equal to $g_+(u_*)$. 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 $g_-(u) = g_+(u)$ for all u versus the general alternative. There is a large literature on testing for structural change in parametric models, Perron (1989), and in nonparametric regression, Muller (1992). Indeed the popular regression discontinuity literature, Imbens and Lemieux (2008), draws on some of these ideas.

4 Estimation

We suppose that a sample $\ell_t, t = 1, \ldots, T$ is observed. 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 $E(\zeta_t^2) <= C < \infty$. In that case, one can estimate the function $g(.)$ by conditional mean smoothing of ℓ_t and the identified parameters β, γ by the GMM approach. Provided that additional high level weak dependence conditions are satisfied, one can ensure a CLT for the resulting estimators. As in Cipollini et al. (2013), one may wish to additionally specify a second conditional moment restriction whereby $E(\zeta_t^2 - (1 + \sigma_\zeta^2)|\mathcal{F}_{t-1}) = 0$ with probability one. This additional moment restriction permits more efficient estimation provided 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 a non-negative i.i.d. sequence, which implies the conditional moment restriction but also many other restrictions. We will consider the mixed continuous/discrete case where for all $x \geq 0$,

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

where f is an absolutely continuous density function with support $(0, \infty)$. In some cases the discrete component (zero returns) is important and in others it is not. For estimation, 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. In that case, the enlarged vector $(\pi, \beta, \gamma, \varphi)^\intercal$ can be estimated by a likelihood method after taking care of $g(.)$. In the case where the density is not parametrically specified, one needs to estimate the error density $f(.)$ along with the trend $g(.)$ and the identified dynamic parameters. For forecasting future values of ℓ_t, one does not need the shock distribution, but prediction intervals and LAR (Liquidity at Risk) requires the estimation of some features of the error distribution.

4.1 Estimation based on conditional moment restriction

We first convert the conditional moment restriction $E(\ell_t|\mathcal{F}_{t-1}) = g(t/T)\lambda_t$ to the unconditional moment restriction

$$E(\ell_t) = g(t/T), \quad t = 1, \ldots, T.$$

We use this condition to obtain an initial consistent estimator of g by the kernel smoothing method, specifically we let

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

where K is a kernel function symmetric about zero supported on $[-1, 1]$ satisfying $\int K(u)du = 1$, while h is a bandwidth sequence. Because of the equally spaced observations in time, the denominator of the Nadaraya-Watson estimator is unnecessary here for interior points (this may also be called the Priestley-Chao estimator). In this context the kernel estimator is Best Linear Minimax (under i.i.d. errors) at any fixed $u \in (0, 1)$ according to Fan (1993). i.e., it is equivalent to the local linear estimator.[1] The estimator does suffer from boundary bias though and in particular $\widehat{g}(0), \widehat{g}(1)$ will not be consistent without modification. A

[1]Notice that provided K and g are twice differentiable with $K(\pm 1) = 0$, $\widehat{g}^{(j)}(u)$ exists and consistently estimates $g^{(j)}(u)$, under some bandwidth conditions, for $j = 1, 2$ and $u \in (0, 1)$.

standard way to correct for boundary bias is to use boundary kernels that adapt to the estimation point as they approach the boundary, Gasser et al. (1985). An alternative method is local linear kernel regression, which does not require an explicit boundary correction, Fan and Gijbels (1996). The issue with both these methods is that the estimate of $g(u)$ is not guaranteed to be nonnegative everywhere, whereas the simple estimator is non-negative with probability one. In practice, at least for our application, this does not seem to be much of an issue and we use also the local linear smoothing method. Nevertheless, for our theoretical results we work with $\widehat{g}(u)$ as above for all $u \in [h, 1-h]$ and in the boundary region we either renormalize by $\sum_{t=1}^{T} K_h(t/T - u)/T$ (standard Nadaraya-Watson estimator) or we set $\widehat{g}(u) = \widehat{g}(h)$ if $u \leq h$ and set $\widehat{g}(u) = \widehat{g}(1-h)$ if $u \geq 1-h$. In this case, we guarantee positivity of our estimate but suffer some performance loss at the boundary. Our estimator of $g(0)$ is consistent provided g is continuous at the boundary because then $g(h) \to g(0)$ as $h \downarrow 0$. Another advantage of the simple estimator is that one can interpret the widely computed measure A_t defined in Equation (1) as a special case of $\widehat{g}(u)$ with uniform kernel and bandwidth equivalent to a month of data around the point u. We define the detrended liquidity $\widehat{\ell}_t^* = \ell_t / \widehat{g}(t/T)$, $t = 1, \ldots, T$.

We next estimate the dynamic parameters $\theta = (\beta, \gamma)^\intercal$ by exploiting the conditional moment restriction

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

where $\ell_t^* = \ell_t / g(t/T)$, $t = 1, \ldots, T$. In other words $\ell_t^* - \lambda_t(\theta)$ is a martingale difference sequence at the true parameter values $\beta = \beta_0, \gamma = \gamma_0$. In practice, we define for any $\theta \in \Theta$, where Θ is a compact set defined below,

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

for $t = 1, \ldots, T$, where we take initializations $\widehat{\ell}_0^*, \lambda_0 = 1$ for simplicity. Then we define $\rho_t(\theta, \widehat{g}) = z_{t-1}(\widehat{\ell}_t^* - \widehat{\lambda}_t(\theta))$, where $z_{t-1} \in \mathcal{F}_{t-1}$ are instruments, and let

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

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

where W is a weighting matrix and $I_T \subset \{1, \ldots, T\}$. In the sequel we suppress the notation I_T, although we discuss this issue in the Appendix. The estimator $\widehat{\theta}_{GMM}$ is consistent for θ under our conditions below (and we drop the subscript GMM below). One can improve the GMM procedure by choosing the instruments and weight matrix optimally, but we shall not pursue this here, since we will pursue efficiency objectives via semiparametric likelihood methods.

We next suggest an improved estimator of $g(.)$. We work from the conditional moment restriction

$$E\left(\frac{\ell_t}{\lambda_t}|\mathcal{F}_{t-1}\right) = g(t/T),$$

which is now feasible given our consistent estimates of θ and hence λ_t. We take a simple implementation of local GMM, Gozalo and Linton (2000) and Lewbel (2007), based only on constant instruments in which case we obtain the closed form

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

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

We have outlined a multistep approach to estimation of $g(.), \theta$. Alternatively, one may use a profile method that links the global and local objective functions. Specifically, for given θ, we estimate $g_\theta(u)$ by smoothing $\ell_t/\lambda_t(\theta, g_\theta)$ against time and optimize the profiled global GMM objective function with respect to θ, letting $\widehat{g}_P(u) = \widehat{g}_{\widehat{\theta}}(u)$. We do not pursue this approach here and focus on multiple step methods.

4.2 Estimation based on i.i.d. assumption

In this case, we assume that the error ζ_t is i.i.d. with mean one, variance σ_ζ^2, and c.d.f. F as specified above. We consider several cases. First, where the density f is known completely. Second, where f is known up to a vector of parameters. Third, where f is of unknown form. Kreiss (1987), Linton (1993), Drost and Klaassen (1997), and Ling and McAleer (2003) consider estimation in dynamic time series models with unknown error density and we follow their approach. In the parametric case, we have a semiparametric model with parameters $(\pi, \beta, \gamma, \varphi^\intercal)^\intercal$ and unknown function $g(.)$. In the case where f is of unknown form, we have a semiparametric model with parameters $(\pi, \beta, \gamma)^\intercal$ and unknown functions $f(.), g(.)$.

For a given density f, define the so-called Fisher location and scale score functions and informations:

$$s_1(\zeta) = -\frac{f'(\zeta)}{f(\zeta)}, \quad s_2(\zeta) = -\left(1 + \zeta\frac{f'(\zeta)}{f(\zeta)}\right) \tag{7}$$

$$I_j(f) = \int s_j^2(\zeta) f(\zeta) d\zeta, \; j = 1, 2. \tag{8}$$

For the Burr density with $f(x; c, k) = ckx^{c-1}/(1+x^c)^{k+1}$, $s_2(\zeta) = -c + (k+1)\zeta^c/(1+\zeta^c)$ is a bounded function of ζ for all parameter values c, k.

4.2.1 Parametric density case

Suppose that f depends on some unknown parameters φ, denoted as f_φ. If $g(.)$ were known, the log likelihood function of $\{\ell_1, \ldots, \ell_T\}$ is, apart from a term to do with $g(.)$ that does not depend on parameters, equal to

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

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

From this we can see the separability of π; the parameter π may be estimated by simply counting the frequency of zeros of ℓ_t.[2] The remaining quantities are estimated using non-zero observations only. To avoid complicating the notation we shall assume in the sequel that $\pi = 0$. In practice, given a consistent estimate of $g(.)$, we may maximize an estimated version of this likelihood $\widehat{L}(\theta, \varphi)$, where $g(.)$ is replaced by $\widehat{g}(.)$ or $\widetilde{g}(.)$. In fact, we avoid further nonlinear optimization by making use of our initial consistent estimates of θ and auxiliary initial consistent estimates of φ, $\widehat{\varphi}$, which may often be obtained through closed form moment estimators. For example, in the gamma case, parameterized to have mean one, the parameter φ can be estimated as one over the variance.

We show in Appendix C.2 of Hafner et al. (2022) that the efficient score functions (in the semiparametric model) for $\eta = (\theta^\intercal, \varphi)^\intercal$ in the presence of unknown $g(.)$ are:

$$L^*_\theta(\eta) = \sum_{t=1}^T \ell^*_{\theta t}(\eta), \quad \ell^*_{\theta t}(\eta) = s_2(\zeta_t)\left(\frac{\partial \log \lambda_t}{\partial \theta} - \frac{E\left[\frac{\partial \log \lambda_t}{\partial \theta}\frac{1}{\lambda_t}\right]}{E\left(\frac{1}{\lambda_t^2}\right)}\frac{1}{\lambda_t}\right) \tag{9}$$

$$L^*_\varphi(\eta) = \sum_{t=1}^T \ell^*_{\varphi t}(\eta), \quad \ell^*_{\varphi t}(\eta) = \frac{\partial \log f_\varphi(\zeta_t)}{\partial \varphi} - \frac{E\left(\frac{\partial \log f_\varphi(\zeta_t)}{\partial \varphi} s_2(\zeta_t)\right) E\left(\frac{1}{\lambda_t}\right)}{I_2(f_\varphi) E\left(\frac{1}{\lambda_t^2}\right)} s_2(\zeta_t)\frac{1}{\lambda_t}. \tag{10}$$

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

$$\widetilde{\eta} = \widehat{\eta} + \mathcal{I}^*_{\eta\eta}(\widehat{\eta}, \widehat{g})^{-1} S^*_\eta(\widehat{\eta}, \widehat{g}), \tag{11}$$

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

[2] This is the approach adopted in Appendix E.8 of Hafner et al. (2022) where we investigate the occurrence of zero returns in the S&P 500 stock market index over the period 1950-2021. The data contains 125 zero returns in total, corresponding to 0.69% of the sample, and the majority of those occurred before 2000.

$$\ell^*_{\theta t}(\widehat{\eta},\widehat{g}) = \widehat{s}_2(\widehat{\zeta}_t)\left(\frac{\partial \log \widehat{\lambda}_t}{\partial\theta} - \frac{\frac{1}{T}\sum_{t=1}^T \frac{\partial \log \widehat{\lambda}_t}{\partial\theta}\frac{1}{\widehat{\lambda}_t}}{\frac{1}{T}\sum_{t=1}^T \frac{1}{\widehat{\lambda}_t^2}}\frac{1}{\widehat{\lambda}_t}\right)$$

$$\ell^*_{\varphi t}(\widehat{\eta},\widehat{g}) = \frac{\partial \log f_{\widehat{\varphi}}\left(\widehat{\zeta}_t\right)}{\partial\varphi} - \frac{\frac{1}{T}\sum_{t=1}^T \frac{\partial \log f_{\widehat{\varphi}}(\widehat{\zeta}_t)}{\partial\varphi}\widehat{s}_2(\widehat{\zeta}_t)\frac{1}{T}\sum_{t=1}^T \frac{1}{\widehat{\lambda}_t}}{\frac{1}{T}\sum_{t=1}^T \widehat{s}_2^2(\widehat{\zeta}_t)\frac{1}{T}\sum_{t=1}^T \frac{1}{\widehat{\lambda}_t^2}}\widehat{s}_2(\widehat{\zeta}_t)\frac{1}{\widehat{\lambda}_t}$$

$$\widehat{s}_2(\zeta) = -\left(1+\zeta\frac{f'_{\widehat{\varphi}}(\zeta)}{f_{\widehat{\varphi}}(\zeta)}\right), \quad \widehat{\lambda}_t = 1-\widehat{\beta}-\widehat{\gamma}+\widehat{\beta}\widehat{\lambda}_{t-1}+\widehat{\gamma}\frac{\ell_{t-1}}{\widehat{g}((t-1)/T)}.$$

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

$$\sum_{t=1}^T K_h(t/T-u)\left(\log f\left(\zeta_t(g)\right) - \log g\right), \tag{12}$$

$$\zeta_t(g) = \frac{\ell_t}{\lambda_t g}, \quad t=1,\ldots,T, \tag{13}$$

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

$$\widetilde{g}_{LocL}(u) = \widehat{g}(u) - \widehat{L}_{gg}^{-1}(\widehat{g}(u);u)\widehat{L}_g(\widehat{g}(u);u), \tag{14}$$

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

$$\widehat{L}(g;u) = \sum_{t=1}^T K_{h^*}(t/T-u)\left(\log f_{\widehat{\varphi}}\left(\widetilde{\zeta}_t(g)\right) - \log g\right) \tag{15}$$

$$\widetilde{\zeta}_t(g) = \frac{\ell_t}{g\lambda_t(\widehat{\theta},\widehat{g})}, \quad t=1,\ldots,T. \tag{16}$$

Here, h^* is a bandwidth sequence.

4.2.2 Nonparametric density case

In Appendix C.3 of Hafner et al. (2022), we derive the efficient score function for θ in the semiparametric model with unknown f, g, thereby extending Drost and Werker (2004). This is

$$L^{**}_\theta(\theta) = \sum_{t=1}^T \ell^{**}_{\theta t}(\theta),$$

$$\ell_{\theta t}^{**}(\theta) = \left(\frac{\zeta_t - 1}{\sigma_\zeta^2} + s_2(\zeta_t)\right) a + s_2(\zeta_t)\left(\frac{\partial \log \lambda_t(\theta)}{\partial \theta} - b\frac{1}{\lambda_t}\right)$$
$$= \frac{\zeta_t - 1}{\sigma_\zeta^2} a + s_2(\zeta_t)\left(\frac{\partial \log \lambda_t(\theta)}{\partial \theta} - a - b\frac{1}{\lambda_t}\right)$$

for some a, b with:

$$a = E\left(\frac{\partial \log \lambda_t}{\partial \theta}\right) - bE\left(\frac{1}{\lambda_t}\right)$$
$$b = \frac{E\left(\frac{1}{\lambda_t}\frac{\partial \log \lambda_t}{\partial \theta}\right) - \kappa E\left(\frac{\partial \log \lambda_t}{\partial \theta}\right) E\left(\frac{1}{\lambda_t}\right)}{E\left(\frac{1}{\lambda_t^2}\right) - \kappa E^2\left(\frac{1}{\lambda_t}\right)},$$

where $\kappa = 1 - 1/I_2(f)\sigma_\zeta^2$.

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

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

where h_f is another bandwidth sequence, and the residuals are defined as

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

This estimator does not impose the restriction that $E(\zeta_t) = 1$, and instead we also consider the estimator based on the rescaled residuals $\widehat{\zeta}_t / \sum_{t=1}^{T} \widehat{\zeta}_t / T$. Notationally, we are assuming the same kernel as in the estimation of the liquidity trend but this need not be the case. In particular, since $\zeta_t \geq 0$ one may wish to use special kernel methods adapted to this problem, Chen (2000) and Scaillet (2004).

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

$$\widetilde{\widehat{\theta}} = \widehat{\theta} + \mathcal{I}_{\theta\theta}^{**}(\widehat{\theta}, \widehat{f}, \widehat{g})^{-1} S_\theta^{**}(\widehat{\theta}, \widehat{f}, \widehat{g}) \tag{17}$$

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

$$\ell_{\theta t}^{**}(\widehat{\theta}, \widehat{f}, \widehat{g}) = \frac{\widehat{\zeta}_t - 1}{\widehat{\sigma}_\zeta^2}\widehat{a} + \widehat{s}_2(\widehat{\zeta}_t)\left(\frac{\partial \log \widehat{\lambda}_t}{\partial \theta} - \widehat{a} - \widehat{b}\frac{1}{\widehat{\lambda}_t}\right)$$

where $\widehat{\sigma}_\zeta^2 = \sum_{t=1}^{T}(\widehat{\zeta}_t - \overline{\widehat{\zeta}})^2/T$ with $\overline{\widehat{\zeta}} = \sum_{t=1}^{T} \widehat{\zeta}_t / T$, and $\widehat{\lambda}_t = \lambda_t(\widehat{\theta}, \widehat{g})$, while

$$\widehat{a} = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \log \widehat{\lambda}_t}{\partial \theta} - \widehat{b}\frac{1}{T}\sum_{t=1}^{T} \frac{1}{\widehat{\lambda}_t}$$

$$\widehat{b} = \frac{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\widehat{\lambda}_t}\frac{\partial \log \widehat{\lambda}_t}{\partial \theta} - \widehat{\kappa}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \log \widehat{\lambda}_t}{\partial \theta}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\widehat{\lambda}_t}}{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\widehat{\lambda}_t^2} - \widehat{\kappa}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\widehat{\lambda}_t}\right)^2},$$

where $\widehat{\kappa} = 1 - 1/I_2(\widehat{f})\widehat{\sigma}_\zeta^2$. Here, $\widehat{1}_t$ is a trimming function that is needed theoretically to reduce the effect of small density estimates. In practice, we have found reasonable performance without trimming. One possible trimming scheme here was considered in Linton and Xiao (2007). In the literature on "adaptive estimation" a number of other devices are used primarily to promote simple proofs, these include discretization of the initial estimator and sample splitting, see for example Kreiss (1987) and Linton (1993).

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

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

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

$$\widehat{L}(g;u) = \sum_{t=1}^{T} K_{h^\dagger}(t/T-u)\left(\log \widehat{f}\left(\widetilde{\zeta}_t(g)\right) - \log g\right), \tag{18}$$

$$\widetilde{\zeta}_t(g) = \frac{\ell_t}{g\widehat{\lambda}_t(\widehat{\theta})}, \quad t = 1,\ldots,T, \tag{19}$$

where $h^\dagger$ is another bandwidth sequence.

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 some defined criterion of convergence. Robinson (1988) establishes some results about the improvements that such iterations can make in theory.

5 Large sample properties

We make some assumptions to support our limiting distributions. The theoretical analysis draws on several previous studies already cited.

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

$$\alpha(k) = \sup_{T\geq 1}\sup_{A\in\mathcal{F}_{-\infty}^{n},\, B\in\mathcal{F}_{n+k}^{\infty}} |P(AB) - P(A)P(B)| \to 0, \tag{20}$$

as $k \to \infty$*, 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 is i.i.d. It may also hold when ζ_t itself is only described as a stationary mixing process although this can be difficult to establish. Instead, one can work with the more general near epoch dependence condition, see Lu and Linton (2007). Recently, Wu (2005) has developed an alternative set of weak dependence conditions that have become very popular and for which many results are available.

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 $\widehat{g}(u)$, $u \in (0,1)$, that is based on smoothing of the raw liquidity. Define the left and right second derivatives of a function $g(.)$ where they exist as

$$g''_+(u) = \lim_{\delta \downarrow 0} \frac{g(u+2\delta) - 2g(u+\delta) + g(u)}{\delta^2}, \; g''_-(u) = \lim_{\delta \uparrow 0} \frac{g(u+2\delta) - 2g(u+\delta) + g(u)}{\delta^2}.$$

The second derivative $g''(u)$ is defined if both $g''_+(u), g''_-(u)$ are defined and are equal.

Assumption A1. *We suppose that $g(.) \in \mathcal{G}$, where for $c > 0$*

$$\mathcal{G} = \left\{ g : g : [0,1] \to \mathbb{R}_+, \; g(x) \geq c, \; \left| g''(x) \right| < \infty \text{ for all } x \in (0,1), \text{ and } g''_+(0), g''_-(1) \text{ exist} \right\}.$$

Define $||g|| = \left(\int_0^1 g(u)^2 du \right)^{1/2}$ and $||g||_\infty = \sup_{u \in [0,1]} |g(u)|$ for all $g \in \mathcal{G}$.

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

$$\alpha(k) \leq C k^{-\rho}.$$

Assumption A3. *Suppose that K is symmetric about zero with compact support $[-1,1]$ such that $K(\pm 1) = 0$ and K is thrice differentiable where K''' is Lipschitz continuous on $[-1,1]$. Let $||K||_2^2 = \int_{-1}^1 K(s)^2 ds$, and $\mu_j(K) = \int_{-1}^1 s^j K(s) ds$, $j = 0,1,2$.*

We may rewrite the model (2), (3) as a nonparametric regression model with trend in mean and variance

$$\ell_t = g(t/T) + g(t/T)v_t,$$

where v_t is a mean zero stationary and alpha mixing series defined in A2. We adapt results of Francisco-Fernández and Vilar-Fernández (2001) for local polynomial estimators in the case without trending heteroskedasticity to obtain the following central limit theorem.

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

$$\sqrt{Th}\left(\widehat{g}(u) - g(u) - h^2 b(u)\right) \Longrightarrow N\left(0, V(u)\right),$$

$$b(u) = \frac{1}{2}\mu_2(K)g''(u)\ ; \quad V(u) = g^2(u) \times ||K||^2 \times \text{lrvar}(v_t).$$

The estimator is consistent and asymptotically normal with optimal rate of $T^{-2/5}$ based on the smoothness assumption. Bandwidth selection procedures and inference procedures require the estimation of $\text{lrvar}(v_t)$, which in general requires further justification. However, we can exploit the fact 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 martingale difference (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 $\widetilde{g}(u)$ 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 $\widehat{\theta}$ is $\sqrt{T}$- consistent. Suppose that $\widetilde{h} = cT^{-1/5}$ for some $c > 0$ and that $Th^5 \to 0$ and $Th/\log T \to \infty$. Then for any $u \in (0,1)$*

$$\sqrt{T\widetilde{h}}\left(\widetilde{g}(u) - g(u) - \widetilde{h}^2 b(u)\right) \Longrightarrow N\left(0, V(u)\right),$$

$$b(u) = \frac{1}{2}\mu_2(K)g''(u)\ ; \quad V(u) = g^2(u)||K||^2 \times \sigma_\zeta^2.$$

The bias term is the same as in Theorem 1 by virtue of the undersmoothing of the first step, see Linton and Xiao (2007). The limiting variance is different though and in particular it is proportional to the variance of ζ_t, which is generally smaller and easier to estimate than the long run variance of v_t. When ζ_t is i.i.d.,

$$E(\lambda_t^2(\zeta_t - 1)^2) = E(\lambda_t^2) \times \sigma_\zeta^2 \geq \sigma_\zeta^2,$$

because $E(\lambda_t^2) \geq 1$ by the Cauchy-Schwarz inequality since $E(\lambda_t) = 1$. For this estimator, consistent standard errors (assuming undersmoothing) can be based on

$$\widehat{V}(u) = \widetilde{g}^2(u)||K||^2 \times \widehat{\sigma}_\zeta^2, \tag{21}$$

where $\widehat{\sigma}_\zeta^2$ is an estimator of σ_ζ^2. The omission of the bias effect in the confidence interval has been subject to criticism and debate and many alternative inference approaches have been suggested, at least in the i.i.d. case, see for example Schennach (2015) and Calonico et al. (2018). One simple approach here is to use the pilot model method used in bandwidth selection, Silverman (1986). Specifically, suppose that $g(u) = \exp(a_0 + a_1 u)$ for some unknown parameters a_0, a_1. In that case, $g''(u) = a_1^2 \exp(a_0 + a_1 u)$ and given estimates of a_0, a_1 (which can be obtained by the OLS of logarithmic liquidity on a constant and trend (with some adjustment)) one can include the estimated bias in the inference procedure.

We may further use this pilot model to select the bandwidth. Since $g''(u)/g(u) = a_1^2$, a "rule-of-thumb" optimal bandwidth procedure here would be

$$h(u) = \left(\frac{||K||_2^2 \widehat{\sigma}_\zeta^2}{\mu_2^2(K) \widehat{a}_1^4} \right)^{1/5} T^{-1/5}, \tag{22}$$

which happens in this case to be constant across $u \in (0, 1)$. Hart (1991) showed that conventional cross-validation fails in settings that include our estimator $\widehat{g}(.)$, that is, the usual recipe for selecting h based on minimizing the leave-one-out (or equivalently penalized) squared residuals will produce $h \sim 0$. This arises because the serial correlation in the error term leads to a bias in the risk estimation. He proposed a modification to address this, that involved estimating the long run variance essentially. We note here that our refined estimator $\widetilde{g}(.)$ is not subject to this criticism, since the error term in that case is a martingale difference sequence. This suggests (although we have not proven this here) that standard cross-validation based on the data $\{\ell_t/\widehat{\lambda}_t\}$ would produce an asymptotically optimal bandwidth choice for $\widetilde{g}(.)$. In summary, the estimator $\widetilde{g}(u)$ has an advantage over $\widehat{g}(u)$ because of the simplicity of handling inference and bandwidth choice questions. On the other hand, the estimator $\widehat{g}(u)$ is a simple linear estimator and is robust to the specification of the process λ_t^*.

5.1.2 Parametric components

Let $\theta = (\beta, \gamma)^\intercal \in \Theta$, where

$$\Theta = \{\theta : \epsilon \leq \beta, \gamma, \beta + \gamma \leq 1 - \epsilon\} \subset \mathbb{R}^2$$

for some $\epsilon > 0$. This guarantees that for example $\lambda_t^*(\theta) \geq \epsilon$ for all $\theta \in \Theta$. We define for any $\theta \in \Theta$ and $g \in \mathcal{G}$

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

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

Assumption A4. *Define:*

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

For all $\delta > 0$, there is an $\epsilon > 0$ such that

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

Uniformly for all $\theta \in \Theta$, the function $M(\theta, g)$ is continuous in g (with respect to the L_2 metric) at $g = g_0$. For all sequences of positive numbers $\delta_T \to 0$:

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

Assumption A5. *The ordinary partial derivative and pathwise derivatives:*

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

$$\Gamma_2(\theta, g_0) \circ (g - g_0) = \frac{\partial}{\partial \tau} M(\theta, g_0 + \tau(g - g_0))\Big|_{\tau=0},$$

are assumed to exist in all directions $\theta \in \Theta_\epsilon \subset \Theta, g \in \mathcal{G}_\epsilon \subset \mathcal{G}$, where $\Theta_\epsilon, \mathcal{G}_\epsilon$ are small neighborhoods of θ_0, g_0 respectively. The matrix $\Gamma(\theta, g_0)$ is continuous in θ at $\theta = \theta_0$ and $\Gamma(\theta_0, g_0)$ is of full rank.

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

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

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

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

Assumption A4 is sufficient for consistency of $\widehat{\theta}$ given that our estimator $\widehat{g}$ is uniformly consistent and $\widehat{g} \in \mathcal{G}$ with probability one. Assumptions A5 and A6 are needed for the asymptotic normality of $\widehat{\theta}$. These conditions have been verified in a number of different model settings under more primitive conditions, see Chen et al. (2003). We establish in our treatment of the properties of $\widehat{g}$ that it is uniformly consistent at a rate that can be better than $T^{-1/4}$, which is also required for this theory. The term $\Gamma_2(\theta, g_0) \circ (g - g_0)$ determines the correction term in the limiting variance and is established in Appendix B of Hafner et al. (2022). We also establish in the appendix that $\sqrt{T}\left(M_T(\theta_0, g_0) + \Gamma_2(\theta_0, g_0) \circ (\widehat{g} - g_0)\right)$ satisfies a CLT, and this as usual requires the undersmoothing of the nonparametric estimation part.

Let

$$w_t = \lambda_t(\zeta_t - 1)z_{t-1} + \frac{1 - \beta - \gamma}{1 - \beta}(\lambda_t\zeta_t - 1)E(z_{t-1}). \tag{23}$$

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

$$\sqrt{T}\left(\widehat{\theta} - \theta\right) \Longrightarrow N(0, V)$$

$$V = \left(\Gamma^{\intercal} W \Gamma\right)^{-1} \left(\Gamma^{\intercal} W \Omega W \Gamma\right) \left(\Gamma^{\intercal} W \Gamma\right)^{-1}, \text{ where } \Omega = \lim_{T\to\infty} \text{var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} w_t\right)$$

In general, the asymptotic variance of $\widehat{\theta}$ will depend on the long run variance of the process w_t and so inference procedures are complicated by that. Nevertheless, standard procedures such as Newey-West can be applied to the residual sequence

$$\widehat{w}_t = \widehat{\lambda}_t(\widehat{\zeta}_t - 1)z_{t-1} + \frac{1 - \widehat{\beta} - \widehat{\gamma}}{1 - \widehat{\beta}}(\widehat{\lambda}_t\widehat{\zeta}_t - 1)\frac{1}{T}\sum_{t=1}^{T} z_t.$$

5.2 Restrictions from i.i.d. shocks

We suppose here that ζ_t is i.i.d. with mean one and density f. We suppose that we have initial consistent estimators of $g(.), \theta$ 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 $\eta = (\theta^{\intercal}, \varphi)^{\intercal}$ and unknown function g. We first consider the local likelihood estimator of the trend function based on the estimated $\widehat{\varphi}$.

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

$$\sqrt{Th_*}\left(\widetilde{g}_{LocL}(u) - g(u) - h_*^2 b(u)\right) \Longrightarrow N\left(0, V(u)\right)$$

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

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

To construct pointwise confidence bands we use:

$$\widetilde{V}(u) = ||K||^2 I_2^{-1}(f_{\widetilde{\varphi}})\widetilde{g}_{LocL}(u)^2,$$

$$I_2(f_{\widetilde{\varphi}}) = \frac{1}{T}\sum_{t=1}^{T}\left(1 + \widetilde{\zeta}_t \frac{f'_{\widetilde{\varphi}}\left(\widetilde{\zeta}_t\right)}{f_{\widetilde{\varphi}}\left(\widetilde{\zeta}_t\right)}\right)^2,$$

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

We next turn to the properties of the estimated parametric components defined in (11). We need some further regularity conditions, basically smoothness and moment conditions about the parametric density function.

Assumption 7. *We suppose that* $\Psi_{k,l}(x;\varphi) = \frac{\partial^{k+l}}{\partial\varphi^k \partial x^l}\log f_\varphi(x)$ *exists and is continuous in both its arguments in a small neighborhood of* φ_0 *and in all* $x \in \mathbb{R}_+$ *for* $k, l = 1, \ldots, 4$ *and that* $\sup_{\varphi:|\varphi-\varphi_0|\leq cT^{-1/2}} |\Psi_{k,l}(\zeta;\varphi) - \Psi_{k,l}(\zeta;\varphi_0)| \leq R(\zeta)$ *for some measurable function* $R(.)$, *where* $E\left(\left(\zeta^l R(\zeta)\right)^\kappa\right), E\left(\left(\zeta^l|\Psi_{k,l}(\zeta;\varphi)|\right)^\kappa\right) < \infty$ *for some* $\kappa \geq 4$ *and for* $l = 0, 1$. *Furthermore, the efficient information matrix*

$$\mathcal{I}^*_{\eta\eta}(\eta) = \begin{pmatrix} \mathcal{I}^*_{\theta\theta} & \mathcal{I}^*_{\theta\varphi} \\ \mathcal{I}^*_{\varphi\theta} & \mathcal{I}^*_{\varphi\varphi} \end{pmatrix} = \begin{pmatrix} E\left(\ell^*_{\theta t}\ell^{*\intercal}_{\theta t}\right) & E\left(\ell^*_{\theta t}\ell^{*\intercal}_{\varphi t}\right) \\ E\left(\ell^*_{\varphi t}\ell^{*\intercal}_{\theta t}\right) & E\left(\ell^*_{\varphi t}\ell^{*\intercal}_{\varphi t}\right) \end{pmatrix},$$

is well defined and positive definite at $\eta = \eta_0$ *and continuous in* η *in a neighborhood of* η_0

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

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

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

$$\mathcal{I}^*_{\eta\eta}(\widetilde{\eta}, \widetilde{g}) = \frac{1}{T}\sum_{t=1}^{T}\ell^*_{\eta t}(\widetilde{\eta}, \widetilde{g})\ell^*_{\eta t}(\widetilde{\eta}, \widetilde{g})^\intercal.$$

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. We turn to the properties of the estimated parametric components defined in (17). We need further regularity conditions and some modifications to our procedure.

Assumption 8. *We suppose that* $f(.)$ *is three times continuously differentiable in* $x \in \mathbb{R}_+$. *Furthermore, the efficient information matrix*

$$\mathcal{I}^{**}_{\theta\theta}(\theta, f, g) = E\left(\ell^{**}_{\theta t}\ell^{**\intercal}_{\theta t}\right).$$

is well defined and positive definite at $\theta = \theta_0$ *and continuous in* θ *in a neighborhood of* θ_0

Theorem 6. *Suppose that Assumptions A1-A6 and A8 hold and that* $\sqrt{T}h^2 \to 0$ *and* $Th/\log T \to \infty$ *and* $\sqrt{T}h_f^2 \to 0$ *and* $Th_f/\log T \to \infty$. *Suppose that* $\widehat{\theta}$ *is* $\sqrt{T}$*-consistent and that certain additional conditions discussed in Appendix B of Hafner et al. (2022) prevail. Then as* $T \to \infty$

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

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

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

6 Testing for temporary and permanent shifts

6.1 Permanent shifts

There is a large literature addressing the question of structural change in nonparametric settings, Muller (1992), Delgado and Hidalgo (2000), Mercurio and Spokoiny (2004), Su and Xiao (2008), and Vogt and Dette (2015). We estimate the function g allowing for a discontinuity at the point $u_* \in (0, 1)$ by considering

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

where K^+ is a kernel supported on $[0, 1]$ with $\int K^+(u)du = 1$ and K^- is a kernel supported on $[-1, 0]$ with $\int K^-(u)du = 1$. We may further require that $\int K^+(u)udu = 0$ and $\int K^-(u)udu = 0$, which improves the magnitude of the bias but has the disadvantage that the corresponding estimators may be negative with positive probability.

We may test for the presence of a discontinuity by computing

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

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

$$\widehat{\sigma}^{2\pm}(u_*) = \widehat{g}^\pm(u_*)^2 \times \widehat{\text{lrvar}}(\lambda_t\zeta_t),$$

because $\ell_t - g(t/T) = g(t/T)\,(\lambda_t\zeta_t - 1)$ has smoothly varying variance that is proportional to $g^2(t/T)$ and because the series $\lambda_t\zeta_t$ is stationary and weakly dependent. On the other hand if we work with the improved estimator that works with a smooth of $\ell_t/\widehat{\lambda}_t$, 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

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

Likewise for the local likelihood estimator, but this also takes care of the error shape and heteroskedasticity when this is correctly specified. Given the studentized statistic, $\tau(u_*)$ 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 $1-\alpha$.

In some cases we may be testing for an effect of an event that takes place at the same time as other structural changes are affecting all stocks. In this case we consider how to include a control group to eliminate common trends at the change time. This amounts to a diff in diff test, Angrist and Pischke (2009). 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 (2) holds for both stocks and that ζ_{St} and ζ_{Ct} may be correlated. We define the diff-in-diff statistic as

$$\tau_{did}(u_*) = \frac{\sqrt{Th}\big(\widetilde{g}^{+}_{S}(u_*)-\widetilde{g}^{-}_{S}(u_*)\big)-\big(\widetilde{g}^{+}_{C}(u_*)-\widetilde{g}^{-}_{C}(u_*)\big)}{\sqrt{\big(\widetilde{\sigma}^{2+}_{S}(u_*)+\widetilde{\sigma}^{2+}_{C}(u_*)-2\widetilde{\sigma}^{+}_{S,C}(u_*)\big)||K^+||^2+\big(\widetilde{\sigma}^{2-}_{S}(u_*)+\widetilde{\sigma}^{2-}_{C}(u_*)-2\widetilde{\sigma}^{-}_{S,C}(u_*)\big)||K^-||^2}},$$

where:

$$\widetilde{\sigma}^{\pm}_{S,C}(u) = \widetilde{g}^{\pm}_{S}(u_*)\widetilde{g}^{\pm}_{C}(u_*) \times \widehat{\sigma}_{\zeta_S,\zeta_C},$$

$$\widehat{\sigma}_{\zeta_S,\zeta_C} = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\zeta}_{St} - \overline{\widehat{\zeta}}_S\right)\left(\widehat{\zeta}_{Ct} - \overline{\widehat{\zeta}}_C\right).$$

This corresponds to a test of the hypothesis that $g^+_S(u_*) - g^-_S(u_*) = g^+_C(u_*) - g^-_C(u_*)$, which imposes weaker assumptions than $g^+_S(u_*) - g^-_S(u_*)$. We comment that the control group approach heavily relies on being able to find stock(s) that are not themselves influenced by the effect on the treatment group, i.e., where spillover effects are not anticipated.

6.2 Temporary effects in the dynamics

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

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

where J is fixed. Here, D_{jt} are dummy variables indicating times $t_1, \ldots, t_J$ and we focus on the case where $t_j = t_1 + j$. 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 $\alpha_1 = \ldots = \alpha_J = 0$ against the general alternative even in the full semiparametric model. The efficient score function with respect to α in the semiparametric model with

parametric f is

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

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

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

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

$$\widehat{CAR}(\tau) = \sum_{j=1}^{\tau} \sum_{t=t_j}^{T} s_2(\widehat{\zeta}_t) \frac{\widehat{\beta}^{t-t_j}}{\lambda_t(\widehat{\theta},0)}, \quad \tau = 1, \ldots, J-1. \tag{24}$$

The test statistics do not satisfy a central limit theorem (even when $\theta, \varphi, g(.)$ are 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 $\theta, g(.)$, we can calculate the distribution numerically using data prior to the event, i.e., $w_r(\tau) = \sum_{j=1}^{\tau}\sum_{t=r}^{T} s_2(\zeta_t)\beta^{t-r}/\lambda_t(\theta,0)$ for r some time before t_1. Let F_w denote the distribution of the series $\{w_r\}$. We assume that there is a long sample of data available before the intervention (i.e., $t_1 \to \infty$ is large) so that this distribution can be consistently estimated from the pre event sample.

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

$$\widehat{w}_r(\tau) = \sum_{j=1}^{\tau} \sum_{t=r}^{T} s_2(\widehat{\zeta}_t) \frac{\widehat{\beta}^{t-r}}{\lambda_t(\widehat{\theta},0)}, \; r = 1, \ldots, t_1 - J. \tag{25}$$

We estimate F_w using the empirical distribution of the data $\{\widehat{w}_r(\tau)\ r = 1, \ldots, t_1 - J\}$.

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

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

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

where R_{mt} and R_{ft} are the market return and risk-free rate respectively. $liq_t^e = E(liq_t|\mathcal{F}_{t-1}) = c_0 + c_1 liq_{t-1}$ 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 $liq_t^u = liq_t - liq_t^e$.

From a time series perspective it seems a bit strange to wait for a whole year to update ones estimate of liquidity. We consider the following specification for daily stock returns

$$E\left(R_{mt} - R_{ft} | \mathcal{F}_{t-1}\right) = a + b \times g(t/T) + c \times \lambda_t + d \times \zeta_t, \tag{26}$$

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, see Escanciano et al. (2017) for related specifications.

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

$$E\left(R_{mt} - R_{ft} - m(t/T) | \mathcal{F}_{t-1}\right) = \alpha + \gamma \times \lambda_t + \delta \times \zeta_t, \tag{27}$$

where $m(t/T) = E\left(R_{mt} - R_{ft}\right)$ is the time varying unconditional equity premium. In practice, we can estimate $m(.)$ by kernel smoothing methods. There is a generated regressor issue here when we replace λ_t and ζ_t by their estimated quantities; we discuss this in Appendix D of Hafner et al. (2022).

8 Empirical study

The ability to accurately model the 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 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.[3] We use historical daily return and volume data, retrieved from Yahoo Finance, to compute the Amihud illiquidity series. 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 1.[4] It can be observed that the Bitcoin asset is less liquid compared

[3]Sample R code can be found at `https://github.com/lw1882/DArLiQ`.

[4]To 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.

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 the others. Moreover, the illiquidity of Facebook stock has higher skewness and thicker tails compared to the other four tech companies.

Table 1: 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 (Appendix E.2 in Hafner et al. (2022)) 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[5]. 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 illiquidity series.[6] 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.

8.1 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, namely the classic specification $\lambda_t = \omega + \beta\lambda_{t-1} + \gamma\ell_{t-1}^*$ and the speci-

[5]We opt for a Gaussian kernel and we choose the bandwidth according to the direct plug-in method as introduced in Ruppert et al. (1995). The magnitude of the bandwidth is in line with the one computed using Equation (22).

[6]Note 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 log illiquidity series as $\log g(t/T)$ is higher than the mean level of $\log \ell_t$ due to Jensen's inequality. In the interest of space, we will focus on the plot of the log illiquidity series hereafter.

fication with asymmetric effect $\lambda_t = \omega + \beta\lambda_{t-1} + \gamma\ell^*_{t-1} + \gamma^-\ell^*_{t-1}\mathrm{I}_{R_{t-1}<0}$. We use a targeting approach in both cases.

8.1.1 Estimation based on conditional moment restrictions

Table 2: Estimated parameters of the λ_t process based on first moment restriction.

	Classic		Asymmetric		
	β	γ	β	γ	$\gamma^{(-)}$
Facebook	0.952	0.024	0.953	0.025	-0.001
	(14.96)	(1.69)	(22.81)	(1.07)	(-0.05)
Amazon	0.947	0.049	0.950	0.038	0.016
	(73.34)	(4.72)	(67.15)	(1.64)	(0.62)
Apple	0.912	0.073	0.911	0.058	0.025
	(53.15)	(6.70)	(63.48)	(2.56)	(0.85)
Google	0.969	0.025	0.966	0.022	0.005
	(56.52)	(2.84)	(63.75)	(2.02)	(0.59)
Microsoft	0.943	0.052	0.937	0.048	0.012
	(78.41)	(5.82)	(94.82)	(4.07)	(0.79)
Bitcoin	0.962	0.030	0.948	0.012	0.029
	(56.13)	(3.23)	(5.96)	(0.11)	(0.15)

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.

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. $E\left[\frac{\ell^*_t}{\lambda_t} - 1 \mid \mathcal{F}_{t-1}\right] = 0$. We further improve the estimates of the $g(t/T)$ function using the estimated $\widehat{\lambda}_t = \widehat{\lambda}_t\left(\widehat{\theta}_{GMM}\right)$ obtained in the previous step. This, in turn, allows us to further improve the estimates of the θ parameters. We report the obtained estimates with associated t-statistics in Table 2. 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 $\widehat{\lambda}_t$ process. The

log transforms of the initial and updated estimates of the trend function, i.e. $\log g(t/T)$, are plotted in Figure 5 (Appendix E.3 of Hafner et al. (2022)) 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.

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

We estimate the model using an alternative approach – the semiparametric MLE approach – where we assume an i.i.d. error term. The conditional distribution of the error term ζ_t can be freely chosen within the class of distributions satisfying the desired requirements, namely the density having non-negative support with unit mean and variance σ_ζ^2. We present estimation results assuming that the error term follows a Weibull$(\Gamma(1+\varphi)^{-1}, \varphi)$ distribution with shape parameter φ. 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. We then obtain the fully efficient estimates with 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 3. The estimates for the parameters of the λ_t process are significant and all illiquidity series exhibit a high degree of persistence as the estimated β coefficients are close to one. In addition, the estimated shape parameters of the Weibull 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 6 and Figure 7 of Appendix E.4 in Hafner et al. (2022) 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 8 and Figure 9 of Appendix E.4 in Hafner et al. (2022). We observe that in most of the cases, there is no evidence suggesting autocorrelation in the residual or squared residual series. Moreover, we use the probability integral transform (PIT) to check how well the assumed Weibull conditional distribution fits the data. The histogram plots of the PITs shown in Figure 10 and Figure 11 (Appendix E.4 of Hafner et al. (2022)) 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. We complement this evidence by estimating the

Table 3: 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	0.859	0.047	1.378	0.735	0.922	-0.028	0.105	1.383	0.732
	(6.39)	(5.13)	(52.76)		(33.83)	(-5.75)	(9.02)	(52.49)	
Amazon	0.916	0.079	1.361	0.743	0.929	0.037	0.065	1.372	0.738
	(287.28)	(27.93)	(103.95)		(352.02)	(13.06)	(21.81)	(103.13)	
Apple	0.879	0.093	1.315	0.768	0.890	0.047	0.077	1.325	0.762
	(291.24)	(51.45)	(950.93)		(319.03)	(38.19)	(27.61)	(1004.26)	
Google	0.909	0.045	1.283	0.786	0.931	-0.012	0.088	1.292	0.781
	(30.88)	(7.75)	(88.69)		(70.45)	(-3.44)	(12.30)	(88.66)	
Microsoft	0.924	0.068	1.377	0.735	0.927	0.039	0.052	1.380	0.733
	(363.95)	(33.80)	(117.78)		(359.73)	(17.89)	(20.26)	(117.78)	
Bitcoin	0.893	0.060	1.137	0.881	0.892	0.040	0.041	1.144	0.876
	(37.33)	(8.36)	(106.49)		(36.69)	(6.61)	(6.65)	(95.55)	

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

tail index of the fitted shock series $\widehat{\zeta}_t$. Our results suggest that the shocks might have a thicker tail than the Weibull distribution while exhibiting under dispersion features.[7] We thus also consider fat-tailed distributions in our analysis, such as the Burr – which nests the Weibull and Lomax distributions as special cases – and Inverse Burr distributions. The estimation results are reported in Appendix E.9.2 of Hafner et al. (2022). We notice that the Lomax distribution provides the worst performance as it lacks the ability to capture the under-dispersion feature with unit mean restriction. The Burr distribution reduces to the Weibull distribution except for Apple and Bitcoin whose shock terms have thicker tails than the other stocks. The Inverse Burr distribution outperforms the Weibull and Burr distributions in terms of log likelihood in most cases but not for Microsoft. No distribution among the ones considered above consistently provides a better fit and we thus refrain from searching for more general distributions. Instead, we will focus in the next section on whether a more flexible nonparametric density can provide a better fit to the data.

Lastly, we further improve the estimation of $g(t/T)$ by maximizing the local likelihood based on the estimated $\widehat{\lambda}_t$ 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 12 (Appendix E.4 in Hafner et al. (2022)) together with the log illiquidity series. As in the GMM case (see Section 8.1.1), we observe that the updated trend function estimates are different from the initial estimate but only to a minor extent.

[7] See Hafner et al. (2022) for more details on the tail index analysis in Appendix E.9.1 and on the magnitude of the error term volatilityTable 5 in Appendix E.9.2.

8.1.3 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 13 and Figure 14 of Appendix E.5 in Hafner et al. (2022), the estimated nonparametric density against the Weibull density using the shape parameter estimates from Section 8.1.2. We observe that the estimated Weibull density curves do not fall between the two standard deviation bands of the estimated nonparametric densities, suggesting that the difference between the estimated parametric and nonparametric densities is statistically significant.

Table 4: Fully efficient estimates of the parameters for λ_t process when using the nonparametric estimates of the density of the error term ζ_t.

	Classic		Asymmetric		
	β	γ	β	γ	$\gamma^{(-)}$
Facebook	0.865	0.050	0.901	-0.017	0.100
	(9.28)	(6.62)	(73.79)	(-4.27)	(10.91)
Amazon	0.916	0.081	0.929	0.039	0.060
	(345.60)	(33.84)	(399.64)	(15.24)	(22.13)
Apple	0.896	0.084	0.908	0.038	0.068
	(186.18)	(29.27)	(245.33)	(15.14)	(21.20)
Google	0.912	0.047	0.926	-0.008	0.084
	(53.12)	(11.26)	(164.03)	(-3.62)	(14.41)
Microsoft	0.929	0.067	0.932	0.040	0.046
	(435.93)	(37.34)	(452.13)	(19.94)	(17.98)
Bitcoin	0.906	0.061	0.901	0.047	0.028
	(54.12)	(10.46)	(70.60)	(9.25)	(4.93)

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.

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 4 and the parameters are all statistically significant. Comparing the estimated values for the λ_t parameters reported in Table 3 and Table 4, 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 5. We conclude that the ML estimation

Table 5: 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	-2171.38	-2139.04	-2135.05	-2091.73	36.32	47.31
Amazon	-4938.91	-4888.77	-4879.22	-4817.83	59.69	70.93
Apple	-8843.96	-8773.77	-8612.49	-8527.32	231.47	246.45
Google	-4019.61	-3967.50	-3953.75	-3872.98	65.86	94.52
Microsoft	-7440.54	-7399.96	-7423.94	-7372.37	16.59	27.59
Bitcoin	-2397.26	-2389.27	-2377.02	-2372.38	20.23	16.89

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.

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.

8.2 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 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 2 the test statistics $\tau(u_0)$ for Apple over its sample period. The plots of the test statistics for the other four tech stocks and Bitcoin asset are presented in Figure 16 and Figure 17 (Appendix E.6 in Hafner et al. (2022)).

We focus on 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 2, Figure 16 and Figure 17. 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 6 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

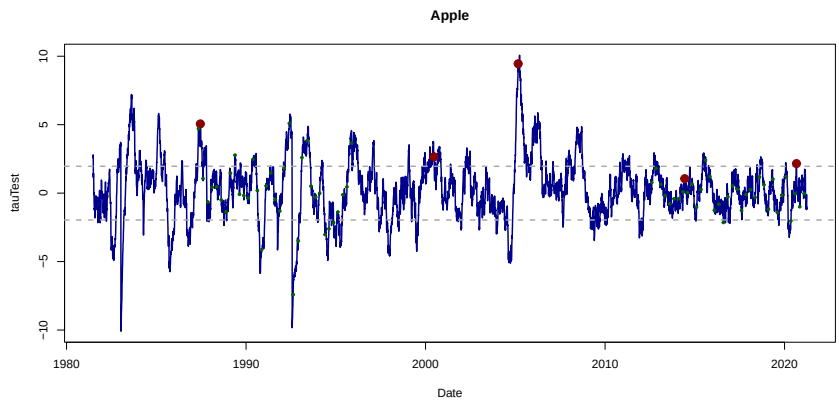


Figure 2: Test statistics for detecting permanent breaks in the illiquidity series.

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.[8]

Table 6: Average statistics for testing permanent breaks in the liquidity series.

	Amazon	Apple	Google	Microsoft	Average
τ	6.507	4.075	3.619	3.113	3.955

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 $\widehat{g}^{\pm}(u)$. Once the detrended illiquidity series are obtained, we use the consistent test developed in Section 6 to test the null hypothesis that $\alpha_1 = \alpha_2 = \ldots = \alpha_J = 0$ against the general alternative semi-parametric 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 7 and Table 8 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.[9] We observe that the effect of stock splits on the short-term dynamics of liquidity is almost always not significant. The only two exceptions

[8]We also consider the diff-in-diff approach where we use Apple or Microsoft as the "control" stock. Results can be found in Table 2 and Table 3 (Appendix E.6 of Hafner et al. (2022)). We find inconclusive evidence for the effect of stock split on illiquidity. This might be due to the fact that the control stock is influenced by spillover effects of the treatment group.

[9]Note 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.

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 suggests 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.

Table 7: Test statistics for detecting permanent and temporary breaks in the liquidity series of Amazon and Apple stocks.

		Split 1	Split 2	Split 3		
Amazon	Date	1998-06-02	1999-01-05	1999-09-02		
	Splits	02:01	03:01	02:01		
	τ_{LR}	3.29	14.67	1.56		
	Δ^1_{LR}	27.71	154.54	13.07		
	Δ^2_{LR}	3.97	22.81	1.10		
	τ_{SR}	17.26	8.09	-20.06		
	$Q^{2.5\%}_{SR}$	-29.75	-42.82	-38.09		
	$Q^{97.5\%}_{SR}$	51.43	57.16	10.15		
		Split 1	Split 2	Split 3	Split 4	Split 5
Apple	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}	5.06	2.65	9.45	1.06	2.16
	Δ^1_{LR}	43.15	22.25	85.76	8.83	18.07
	Δ^2_{LR}	0.10	0.10	0.68	0.04	0.40
	τ_{SR}	2.97	10.45	-1.54	5.14	13.45
	$Q^{2.5\%}_{SR}$	-18.51	-15.74	-11.69	-13.85	-10.77
	$Q^{97.5\%}_{SR}$	24.43	20.54	20.06	14.08	15.95

Note: We report the test statistic for permanent breaks (τ_{LR}), the percentage change in the long-run illiquidity level (Δ^1_{LR}) and the variation in the long-run illiquidity level normalized by the change in the real tick size ($\Delta^2_{LR} \times 10^8$). We also report the test statistic values for the temporary shifts (τ_{SR}) together with the 2.5% and 97.5% quantiles of τ_{SR} which are estimated based on past data.

We further investigate the economic magnitude of the permanent effect of stock splits on liquidity conditions. In particular, we consider two statistics: i) the percentage change in the long-run illiquidity level, i.e. $\Delta^1_{LR} = \frac{2\left(g^+(u)-g^-(u)\right)}{g^+(u)+g^-(u)}$; ii) the variation in the long-run illiquidity level normalized by the change in the real tick size, i.e. $\Delta^2_{LR} = \frac{g^+(u)-g^-(u)}{\frac{\kappa(u)}{P(u)} - \frac{\kappa(u)}{sP(u)}}$ where κ represents the tick size, P is the stock price and s represents the split factor. Based on the statistics reported in Table 7 and Table 8, the 1999 Amazon stock split has the largest impact on liquidity in terms of percentage change in the long-run illiquidity level (Δ^1_{LR}) and changes in the long-run illiquidity level normalized by the real tick size. However, the 2014 Google stock split has the largest impact when considering the measure normalized by the change in real tick size (Δ^2_{LR}). In addition, we notice that despite Apple and Microsoft having larger amount of stock splits compared to the other companies analyzed, the majority of them did not have an economically significant impact on liquidity – especially when normalizing by the real tick size.

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

		Split 1				
Google	Date	2014-03-27				
	Splits	2002:1000				
	τ_{LR}	7.04				
	Δ^1_{LR}	69.78				
	Δ^2_{LR}	57.24				
	τ_{SR}	-1.05				
	$Q^{2.5\%}_{SR}$	-22.15				
	$Q^{97.5\%}_{SR}$	23.75				
		Split 1	Split 2	Split 3	Split 4	Split 5
Microsoft	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}	9.71	4.33	2.08	2.14	1.59
	Δ^1_{LR}	67.19	28.71	13.67	14.06	10.43
	Δ^2_{LR}	0.82	0.15	0.10	0.11	0.04
	τ_{SR}	10.71	1.42	31.73	8.08	-3.29
	$Q^{2.5\%}_{SR}$	-31.86	-21.37	-14.26	-16.86	-17.32
	$Q^{97.5\%}_{SR}$	27.56	26.12	26.22	18.78	22.34
		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}	2.13	3.32	2.46	0.25	
	Δ^1_{LR}	14.03	21.88	16.18	1.62	
	Δ^2_{LR}	0.00	0.21	0.23	0.07	
	τ_{SR}	28.48	6.04	9.88	13.95	
	$Q^{2.5\%}_{SR}$	-15.23	-14.65	-15.80	-19.89	
	$Q^{97.5\%}_{SR}$	19.38	18.26	16.32	27.54	

Note: We report the test statistic for permanent breaks (τ_{LR}), the percentage change in the long-run illiquidity level (Δ^1_{LR}) and the variation in the long-run illiquidity level normalized by the change in the real tick size ($\Delta^2_{LR} \times 10^8$). We also report the test statistic values for the temporary shifts (τ_{SR}) together with the 2.5% and 97.5% quantiles of τ_{SR} which are estimated based on past data.

8.3 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 illiquidty 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.7 in Hafner et al. (2022). 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.[10] 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.

Table 9: Coefficient estimates for regressions using daily, weekly and monthly observations.

	Daily	Weekly	Monthly
α	-0.0007	-0.0043	-0.0505^{**}
	(0.0005)	(0.0032)	(0.0168)
γ	0.0017^{***}	0.0065^{*}	0.0556^{**}
	(0.0004)	(0.0031)	(0.0168)
δ	-0.0010^{***}	-0.0023^{***}	-0.0046^{*}
	(0.0001)	(0.0006)	(0.0022)
R^2	0.0068	0.0074	0.0200
Adj. R^2	0.0066	0.0067	0.0174
Num. obs.	15387	2893	741

Note: We estimate the regression based on Equation (27): $R_{mt} - R_{ft} - m(t/T) = \alpha + \gamma \times \lambda_t + \delta \times \zeta_t + \varepsilon_t$, where $m(t/T)$ is the time-varying unconditional equity premium. The significance level is indicated by $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

We consider the specification from Equation (27) in Section 7 for the regression of detrended risk premium on illiquidity components.[11] The estimation results for the three sampling frequencies considered are provided in Table 9. 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 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.

9 Concluding remarks

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

[10] This is confirmed by regression results based on Equation (26) introduced in Section 7. The coefficient estimates for the parameter b associated with the long-run trend illiquidity component are not significant.

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

variations, which makes the stationarity assumption unreasonable for illiquidity modelling. We developed a class of dynamic semiparametric models that captures long-term trend in a flexible way and short-run variations with an autoregressive component. Our 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. The assumption of multiplicative trend is consistent with the general modelling strategy we have adopted but we may also consider a more general model in which everything changes over time, the so-called TV-MEM model. The TV-GARCH model has been investigated in Rohan and Ramanathan (2013) following Dahlhaus and Rao (2006). We also plan to extend our work to cover the multiviarate case.

We further developed a methodology to detect the occurrence of permanent and temporary shifts in the illiquidity process at a given point in time. We applied this framework to study how stock splits affect liquidity. 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).

We also investigated 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.

Acknowledgements

We are grateful for discussions with seminar participants at NYU Abu Dhabi, University of York, Humboldt-Universität zu Berlin, the Janeway Institute Cambridge and the conference participants at the 2022 QFFE in Marseille. We would also like to thank Andrew Harvey for helpful comments. The first author acknowledges financial support of the ARC research contract 18/23-089 of the Belgian Federal Science Policy Office. The second and third authors acknowledge financial support from the Janeway Institute.

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.

Angrist, J. D. and Pischke, J.-S. (2009): *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

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

Beddington, J., Furse, C., Bond, P., Cliff, D., Goodhart, C., Houstoun, K., Linton, O. B., 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.

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

Bickel, P. J., Klaassen, C. A., Ritov, Y. and Wellner, J. A. (1993): *Efficient and adaptive estimation for semiparametric models*. Vol. 4. Springer.

Brogaard, J. (2010): "High frequency trading and its impact on market quality". *Northwestern University Kellogg School of Management Working Paper* 66.

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

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

Chen, X. (2007): "Large sample sieve estimation of semi-nonparametric models". *Handbook of Econometrics* 6, pp. 5549–5632.

Chen, X., Linton, O. B. and Van Keilegom, I. (2003): "Estimation of semiparametric models when the criterion function is not smooth". *Econometrica* 71.5, pp. 1591–1608.

Cipollini, F., Engle, R. F. and Gallo, G. M. (2013): "Semiparametric vector MEM". *Journal of Applied Econometrics* 28.7, pp. 1067–1086.

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

Copeland, T. E. (1979): "Liquidity changes following stock splits". *The Journal of Finance* 34.1, pp. 115–141.

Dahlhaus, R. and Rao, S. S. (2006): "Statistical inference for time-varying ARCH processes". *The Annals of Statistics* 34.3, pp. 1075–1114.

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Hafner, C., Linton, O. B. and Wang, L. (2022): "Supplementary Material for 'Dynamic Autoregressive Liquidity (DArLiQ)'".

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

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

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

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

Imbens, G. W. and Lemieux, T. (2008): "Regression discontinuity designs: A guide to practice". *Journal of Econometrics* 142.2, pp. 615–635.

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

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

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

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

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

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

Lu, Z. and Linton, O. B. (2007): "Local linear fitting under near epoch dependence". *Econometric Theory* 23.1, pp. 37–70.

Mercurio, D. and Spokoiny, V. (2004): "Statistical inference for time-inhomogeneous volatility models". *The Annals of Statistics* 32.2, pp. 577–602.

Muller, H.-G. (1992): "Change-points in nonparametric regression analysis". *The Annals of Statistics* 20.2, pp. 737–761.

Nelson, D. B. and Cao, C. Q. (1992): "Inequality constraints in the univariate GARCH model". *Journal of Business & Economic Statistics* 10.2, pp. 229–235.

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

Perron, P. (1989): "The great crash, the oil price shock, and the unit root hypothesis". *Econometrica* 57.6, pp. 1361–1401.

Robinson, P. M. (1988): "The stochastic difference between econometric statistics". *Econometrica* 56.3, pp. 531–548.

Rohan, N. and Ramanathan, T. (2013): "Nonparametric estimation of a time-varying GARCH model". *Journal of Nonparametric Statistics* 25.1, pp. 33–52.

Ruppert, D., Sheather, S. J. and Wand, M. P. (1995): "An effective bandwidth selector for local least squares regression". *Journal of the American Statistical Association* 90.432, pp. 1257–1270.

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

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

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

Su, L. and Xiao, Z. (2008): "Testing structural change in time-series nonparametric regression models". *Statistics and its Interface* 1.2, p. 347.

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

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

Vogt, M. and Dette, H. (2015): "Detecting gradual changes in locally stationary processes". *The Annals of Statistics* 43.2, pp. 713–740.

Wu, W. B. (2005): "Nonlinear system theory: Another look at dependence". *Proceedings of the National Academy of Sciences* 102.40, pp. 14150–14154.

Supplementary Material for "Dynamic Autoregressive Liquidity (DArLiQ)"

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

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

[b]Faculty of Economics, University of Cambridge

September 16, 2022

Appendices

A Lemmas

Lemma 1. *Suppose that Assumptions A1-A3 hold. Then, we have for any u*

$$\widehat{g}(u) - g(u) = V_T(u) + B_T(u) + R_T(u),$$

where $B_T(u)$ is deterministic and

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

where:

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

$$\sup_{u\in[h,1-h]} \left|R_T(u)\right| = o_P(h^2).$$

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

[†]Email: obl20@cam.ac.uk.

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

$$\sup_{u\in[0,h]\cup[1-h,1]} \left|B_T(u)\right| = O(h), \qquad \sup_{u\in[0,h]\cup[1-h,1]} \left|R_T(u)\right| = o_P(h).$$

Proof of Lemma 1. We write

$$\widehat{g}(u) - g(u) = \frac{1}{T}\sum_{t=1}^{T} K_h(t/T-u)g(t/T)v_t + \frac{1}{T}\sum_{t=1}^{T} K_h(t/T-u)g(t/T) - g(u). \tag{1}$$

Write

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} K_h(t/T-u)g(t/T)v_t &= g(u)\frac{1}{T}\sum_{t=1}^{T} K_h(t/T-u)v_t + \frac{1}{T}\sum_{t=1}^{T} K_h(t/T-u)\left(g(t/T)-g(u)\right)v_t \\ &= V_{T1}(u) + V_{T2}(u).\end{aligned}$$

Furthermore,

$$\sup_{u\in[0,1]} \left|V_{T1}(u)\right| \leq \sup_{u\in[0,1]} g(u) \times \sup_{u\in[0,1]} \left|\frac{1}{T}\sum_{t=1}^{T} K_h(t/T-u)v_t\right| = O_P\left(\sqrt{\frac{\log T}{Th}}\right),$$

by standard arguments applied to $\sum_{t=1}^{T} K_h(t/T-u)v_t$ since v_t is assumed to be stationary and mixing, Francisco-Fernández et al. (2003). We have by Taylor expansion

$$\begin{aligned}V_{T2}(u) &= hg'(u)\frac{1}{T}\sum_{t=1}^{T} L_{1h}(t/T-u)v_t + h^2 g''(u)\frac{1}{2T}\sum_{t=1}^{T} L_{2h}(t/T-u)v_t \\ &\quad + h^2\frac{1}{2T}\sum_{t=1}^{T} L_{2h}(t/T-u)\left(g''(q^*(t/T,u)) - g''(u)\right)v_t,\end{aligned}$$

where $L_j(v) = K(v)v^j$, $j = 1,2$ and $q^*(t/T,u)$ is an intermediate point. By the same type of arguments $\sum_{t=1}^{T} L_{jh}(t/T-u)v_t/T = O_P\left(\sqrt{\frac{\log T}{Th}}\right)$. The last term is $o_P(h^2)$ by the continuity of $g''(.)$ and the fact that

$$\sup_{u\in[0,1]} \frac{1}{T}\sum_{t=1}^{T} \left|L_{2h}(t/T-u)v_t\right| = O_P(1).$$

The bias approximation is valid over $[h, 1-h]$ by standard Taylor series argument. Furthermore, since $g(h) - g(\theta h) = (1-\theta)hg'(\theta h) + O(h^2)$ the approximation over $[0,h]$ is valid, likewise for $[1-h,1]$.

Futhermore,

$$\Pr\left(\min_{1\leq t\leq T} \widehat{g}(t/T) < c/2\right) \longrightarrow 0.$$

This follows because $\mathcal{A} = \left\{\min_{1\leq t\leq T} \widehat{g}(t/T) < c/2\right\} \subset \mathcal{B} = \left\{\max_{1\leq t\leq T} \left|\widehat{g}(t/T) - g(t/T)\right| > c/2\right\}$, where $\Pr(\mathcal{B}) \to 0$ by the uniform expansion in Lemma 1.

Define the infeasible estimators based on the iid sequence ζ_t whose density is f supported on $\mathbb{R}_+$,

$$\widetilde{f}(\zeta) = \frac{1}{T}\sum_{t=1}^{T} K_{h_f}\left(\zeta_t - \zeta\right), \quad \widetilde{s}_2(\zeta) = -\left(\zeta\frac{\widetilde{f}'(\zeta)}{\widetilde{f}(\zeta)} + 1\right).$$

We have the standard result under our conditions.

Lemma 2. *We have*

$$\sup_{\zeta\in\mathbb{R}_+}\left|\widetilde{f}(\zeta) - E(\widetilde{f}(\zeta))\right| = O_P\left(\sqrt{\frac{\log T}{Th_f}}\right)$$

$$\sup_{\zeta\in\mathbb{R}_+}\left|\widetilde{f}'(\zeta) - E(\widetilde{f}'(\zeta))\right| = O_P\left(\sqrt{\frac{\log T}{Th_f^3}}\right).$$

Furthermore, for some sequences $c_{1T} \to 0$ and $c_{2T} \to \infty$

$$\sup_{c_{1T}\le\zeta\le c_{2T}}\left|\widetilde{s}_2(\zeta) - s_2(\zeta)\right| = O_P\left(\sqrt{\frac{\log T}{Th_f^3}} + h_f^2\right).$$

The sequence c_{2T} is needed because $f(\zeta) \to 0$ as $\zeta \to \infty$, the sequence c_{1T} is needed because of boundary issues for the bias terms. The proofs of these results are standard and ommitted. We also have the following result for the feasible density estimator.

Lemma 3. *For some sequences $c_{1T} \to 0$ and $c_{2T} \to \infty$*, we *have*

$$\sup_{c_{1T}\le\zeta\le c_{2T}}\left|\widehat{f}(\zeta) - f(\zeta)\right| = O_P\left(\sqrt{\frac{\log T}{Th_f}} + h_f^2 + h^2\right)$$

$$\sup_{c_{1T}\le\zeta\le c_{2T}}\left|\widehat{f}'(\zeta) - f(\zeta)\right| = O_P\left(\sqrt{\frac{\log T}{Th_f^3}} + h_f^2 + h^2\right).$$

$$\sup_{c_{1T}\le\zeta\le c_{2T}}\left|\widehat{s}_2(\zeta) - s_2(\zeta)\right| = O_P\left(\sqrt{\frac{\log T}{Th_f^3}} + h_f^2 + h^2\right).$$

Proof of Lemma 3. We have

$$\widehat{f}(\zeta) - \widetilde{f}(\zeta) = \frac{1}{Th_f^2}\sum_{t=1}^{T} K'\left(\frac{\zeta_t - \zeta}{h_f}\right)\left(\widehat{\zeta}_t - \zeta_t\right) + \frac{1}{2Th_f^3}\sum_{t=1}^{T} K''\left(\frac{\zeta_t - \zeta}{h_f}\right)\left(\widehat{\zeta}_t - \zeta_t\right)^2$$
$$\frac{1}{2Th_f^3}\sum_{t=1}^{T}\left(K''\left(\frac{\overline{\zeta}_t - \zeta}{h_f}\right) - K''\left(\frac{\zeta_t - \zeta}{h_f}\right)\right)\left(\widehat{\zeta}_t - \zeta_t\right)^2,$$

$$\widehat{f}'(\zeta) - \widetilde{f}'(\zeta) = \frac{1}{Th_f^2}\sum_{t=1}^{T} K''\left(\frac{\zeta_t - \zeta}{h_f}\right)\left(\widehat{\zeta}_t - \zeta_t\right) + \frac{1}{2Th_f^3}\sum_{t=1}^{T} K'''\left(\frac{\zeta_t - \zeta}{h_f}\right)\left(\widehat{\zeta}_t - \zeta_t\right)^2$$
$$\frac{1}{2Th_f^3}\sum_{t=1}^{T}\left(K'''\left(\frac{\overline{\zeta}_t - \zeta}{h_f}\right) - K'''\left(\frac{\zeta_t - \zeta}{h_f}\right)\right)\left(\widehat{\zeta}_t - \zeta_t\right)^2,$$

where $\overline{\zeta}_t$ is an intermediate point. We next substitute in the expansion (5) for $\widehat{\zeta}_t - \zeta_t$ and work term by term. The remainder term uses the Lipschitz continuity of K''' and the uniform convergence rate of $\widehat{\zeta}_t - \zeta_t$.

B Proof of main results

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

$$\sqrt{Th}V_T(u) \Longrightarrow N(0, ||K||_2^2 g(u)^2 \text{lrvar}(v_t)),$$

by the arguments of Francisco-Fernández and Vilar-Fernández (2001) based on the CLT for mixing random variables.

Proof of Theorem 2. First, note that

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T - u)\frac{\ell_t}{\lambda_t} - g(u) &= \frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T - u)g(t/T)\zeta_t - g(u) \\
&= g(u)\frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T - u)\left(\zeta_t - 1\right) \\
&+ \frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T - u)g(t/T) - g(u) \\
&+ \frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T - u)\left(g(t/T) - g(u)\right)\left(\zeta_t - 1\right) \\
&= V_T^+(u) + B_T^+(u) + R_T^+(u),
\end{aligned}$$

where $V_T^+(u)$ is a mean zero stochastic term, whereas $B_T^+(u) = B_T(u)$ is the deterministic bias term, while $R_T^+(u) = o_P(\widetilde{h}^2)$. The term $V_T^+(u)$ has a MDS error term and satisfies the CLT

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

We next show that this is the leading term.

We have

$$\lambda_t(\widehat{\theta}, \widehat{g}) - \lambda_t = \lambda_t(\widehat{\theta}, g_0) - \lambda_t + \lambda_t(\theta_0, \widehat{g}) - \lambda_t + Rem_{t,T},$$

where the remainder term $Rem_{t,T}$ is of smaller order. We focus on the two "linear" terms. We have

$$\lambda_t(\widehat{\theta}, g_0) - \lambda_t = \frac{\partial \lambda_t(\theta_0, g_0)}{\partial \theta^\intercal}\left(\widehat{\theta} - \theta_0\right) + \left(\widehat{\theta} - \theta_0\right)^\intercal \frac{\partial^2 \lambda_t(\theta_0, g_0)}{\partial \theta \partial \theta^\intercal}\left(\widehat{\theta} - \theta_0\right) \tag{2}$$
$$+ \left(\widehat{\theta} - \theta_0\right)^\intercal \left(\frac{\partial^2 \lambda_t(\overline{\theta}, g_0)}{\partial \theta \partial \theta^\intercal} - \frac{\partial^2 \lambda_t(\overline{\theta}, g_0)}{\partial \theta \partial \theta^\intercal}\right)\left(\widehat{\theta} - \theta_0\right),$$

where $\left\|\overline{\theta} - \theta_0\right\| \leq \left\|\widehat{\theta} - \theta_0\right\|$. We have, ignoring initial conditions

$$\begin{aligned}
\lambda_t(\theta_0, \widehat{g}) - \lambda_t &= \gamma_0 \sum_{j=1}^{t} \beta_0^{j-1}\left(\frac{\ell_{t-j}}{\widehat{g}((t-j)/T)} - \frac{\ell_{t-j}}{g((t-j)/T)}\right) \\
&= -\gamma_0 \sum_{j=1}^{t} \beta_0^{j-1} \frac{\ell_{t-j}}{g((t-j)/T)} \frac{\widehat{g}((t-j)/T) - g((t-j)/T)}{g((t-j)/T)} \\
&+ \frac{\gamma_0}{2} \sum_{j=1}^{t} \beta_0^{j-1} \frac{\ell_{t-j}}{g((t-j)/T)} \left(\frac{\widehat{g}((t-j)/T) - g((t-j)/T)}{g((t-j)/T)}\right)^2 + Rem_{t,T},
\end{aligned}$$

where the remainder term $Rem_{t,T}$ is of smaller order.

We have

$$\begin{aligned}
R_T^* &= \left|\frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T - u)\frac{\ell_t}{\widehat{\lambda}_t} - \frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T - u)\frac{\ell_t}{\lambda_t}\right| \\
&= \left|-\frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T - u)\frac{\ell_t}{\lambda_t}\frac{\lambda_t(\widehat{\theta}, \widehat{g}) - \lambda_t}{\lambda_t(\widehat{\theta}, \widehat{g})}\right| \\
&\leq \max_{1 \leq t \leq T}\left|\frac{1}{\lambda_t(\widehat{\theta}, \widehat{g})}\right| \left|\frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T - u)\frac{\ell_t}{\lambda_t}\left(\lambda_t(\widehat{\theta}, \widehat{g}) - \lambda_t\right)\right| \\
&\leq O_P(1) \times \left|\frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T - u)\frac{\ell_t}{\lambda_t}\left(\lambda_t(\widehat{\theta}, \widehat{g}) - \lambda_t\right)\right|,
\end{aligned}$$

because $\lambda_t(\theta, g) \geq \epsilon$ for all $\theta \in \Theta$ and $g \in \mathcal{G}$, and indeed

$$\lambda_t(\widehat{\theta}, \widehat{g}) = \lambda_t(\theta_0, g_0) - \left|\lambda_t(\widehat{\theta}, \widehat{g}) - \lambda_t(\theta_0, g_0)\right| \geq \lambda_t(\theta_0, g_0) - o_P(1),$$

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

We have

$$\frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T - u)\frac{\ell_t}{\lambda_t}\frac{\partial \lambda_t(\theta_0, g_0)}{\partial \theta^\intercal}\left(\widehat{\theta} - \theta_0\right) = O_P(T^{-1/2}).$$

We next consider the nonparametric part:

$$
\begin{aligned}
S_T &= \frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T-u)\frac{\ell_t}{\lambda_t}\sum_{j=1}^{t}\beta_0^{j-1}\frac{\ell_{t-j}}{g((t-j)/T)}\frac{\widehat{g}((t-j)/T)-g((t-j)/T)}{g((t-j)/T)} \\
&= \frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T-u)\frac{\ell_t}{\lambda_t}\sum_{j=1}^{t}\beta_0^{j-1}\frac{\ell_{t-j}}{g((t-j)/T)}\frac{V_T((t-j)/T)}{g((t-j)/T)} \\
&+ \frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T-u)\frac{\ell_t}{\lambda_t}\sum_{j=1}^{t}\beta_0^{j-1}\frac{\ell_{t-j}}{g((t-j)/T)}\frac{B_T((t-j)/T)}{g((t-j)/T)} \\
&+ \frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T-u)\frac{\ell_t}{\lambda_t}\sum_{j=1}^{t}\beta_0^{j-1}\frac{\ell_{t-j}}{g((t-j)/T)}\frac{R_T((t-j)/T)}{g((t-j)/T)} \\
&= S_{T1}+S_{T2}+S_{T3}.
\end{aligned}
$$

Clearly, $S_{T2} = O_P(h^2) = o_P(\widetilde{h}^2)$, $S_{T_3} = o_P(h^2) = o_P(\widetilde{h}^2)$ by the undermsoothing, so we consider S_{T1}, which is

$$
S_{T1} = \frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T-u)\frac{\ell_t}{\lambda_t}\sum_{j=1}^{t}\beta_0^{j-1}\frac{\ell_{t-j}}{g((t-j)/T)}\frac{g(u)}{g((t-j)/T)}\frac{1}{T}\sum_{s=1}^{T} K_h(s/T-(t-j)/T)v_s.
$$

Consider

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T-u)\frac{\ell_t}{\lambda_t}\frac{\ell_{t-1}}{g((t-1)/T)}\frac{g(u)}{g((t-1)/T)}\frac{1}{T}\sum_{s=1}^{T} K_h(s/T-(t-1)/T)v_s \\
&\simeq g(u)\frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T-u)\zeta_t\lambda_{t-1}\zeta_{t-1}\frac{1}{T}\sum_{s=1}^{T} K_h(s/T-(t-1)/T)v_s \\
&= g(u)E\left(\zeta_t\lambda_{t-1}\zeta_{t-1}\right)\frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T-u)\frac{1}{T}\sum_{s=1}^{T} K_h(s/T-(t-1)/T)v_s \\
&+ g(u)\frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T-u)\left(\zeta_t\lambda_{t-1}\zeta_{t-1}-E\left(\zeta_t\lambda_{t-1}\zeta_{t-1}\right)\right)\frac{1}{T}\sum_{s=1}^{T} K_h(s/T-(t-1)/T)v_s.
\end{aligned}
$$

We have

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} K_{\widetilde{h}}(t/T-u)\frac{1}{T}\sum_{s=1}^{T} K_h(s/T-(t-1)/T)v_s \\
&\simeq \frac{1}{T^2}\sum_{s=1}^{T}\left(\sum_{t=1}^{T} K_{\widetilde{h}}(t/T-u)K_h(s/T-t/T)\right)v_s,
\end{aligned}
$$

which is mean zero and has variance

$$
\frac{1}{T^4}\sum_{s=1}^{T}\sum_{s'=1}^{T}\kappa_{Ts}\kappa_{Ts'}E\left(v_s v_{s'}\right),
$$

where $\kappa_{Ts} = \sum_{t=1}^{T} K_{\widetilde{h}}(t/T-u)K_h(s/T-t/T)$. We have

$$\frac{1}{T^4}\sum_{s=1}^{T}\kappa_{Ts}^2 = \frac{1}{T^4}\frac{1}{\widetilde{h}^2h^2}\sum_{s=1}^{T}\left(\sum_{t=1}^{T}K\left(\frac{t/T-u}{\widetilde{h}}\right)K\left(\frac{s/T-t/T}{h}\right)\right)^2$$
$$= \frac{1}{T^4}\frac{1}{\widetilde{h}^2h^2}\sum_{s=1}^{T}\sum_{t=1}^{T}K\left(\frac{t/T-u}{\widetilde{h}}\right)^2K\left(\frac{s/T-t/T}{h}\right)^2 +$$
$$+\frac{1}{T^4}\frac{1}{\widetilde{h}^2h^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{t'=1}^{T}K\left(\frac{t/T-u}{\widetilde{h}}\right)K\left(\frac{s/T-t/T}{h}\right)K\left(\frac{t'/T-u}{\widetilde{h}}\right)K\left(\frac{s/T-t'/T}{h}\right)$$

We have

$$\left\{(t,s)\in\{1,\ldots,T\}^2 : \left|\frac{t}{T}-u\right|\le\widetilde{h}, \left|\frac{s}{T}-\frac{t}{T}\right|\le h\right\} = O(Th)+O(T\widetilde{h})$$

$$\left\{(t',t,s)\in\{1,\ldots,T\}^3 : \left|\frac{t}{T}-u\right|\le\widetilde{h}, \left|\frac{t'}{T}-u\right|\le\widetilde{h}, \left|\frac{s}{T}-\frac{t}{T}\right|\le h, \left|\frac{s}{T}-\frac{t'}{T}\right|\le h\right\}$$
$$= O((Th)^2)+O((T\widetilde{h})^2)$$

Therefore

$$\frac{1}{T^4}\frac{1}{\widetilde{h}^2h^2}\times\left(O((Th)^2)+O((T\widetilde{h})^2)\right) = O\left(\frac{1}{T^2h^2}+\frac{1}{T^2\widetilde{h}^2}\right).$$

It follows that

$$S_{T1} = O_P\left(T^{-1}\widetilde{h}^{-1}\right) = o_P(T^{-1/2}h^{-1/2})$$
$$R_T^* = o_P(T^{-1/2}h^{-1/2})+o_P(h^2).$$

Proof of Theorem 3. We apply Theorem 1 and 2 of Chen et al. (2003). We note that Lemma 1 establishes that

$$\sup_{u\in[h,1-h]}\left|\widehat{g}(u)-g(u)\right| = o_P(T^{-1/4}).$$

We note that this is all that is required since one can drop from the calculation of M_T the observations $t=1,\ldots,Th$ and $t=T-Th,\ldots,T$, that is, by taking $I_T=\{t: Th+1,\ldots,T-Th\}$.

We next establish that

$$\sqrt{T}\left(M_T(\theta_0,g_0)+\Gamma_2(\theta_0,g_0)\circ(\widehat{g}-g_0)\right)\Longrightarrow N(0,\Omega).$$

In the sequel we proceed to infinity to simplify the presentation. We consider

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

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

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

We next calculate

$$\frac{\partial}{\partial\tau}M(\theta, g_0 + \tau(g - g_0)). \tag{3}$$

We have for any τ

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

and so

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

Furthermore, for

$$\frac{\frac{\ell_t}{g_0+\tau(g-g_0)(t/T)} - \frac{\ell_t}{g_0(t/T)}}{\tau} \simeq -\frac{\ell_t}{g_0(t/T)}\frac{g(t/T) - g_0(t/T)}{g_0(t/T)}$$

$$E\left(\frac{\frac{\ell_t}{g_0+\tau(g-g_0)} - \frac{\ell_t}{g_0(t/T)}}{\tau}\right) \simeq -\frac{g(t/T) - g_0(t/T)}{g_0(t/T)}.$$

Therefore,

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

We have

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

It follows that

$$M_T(\theta, \widehat{g}) = \frac{1}{T}\sum_{t=1}^{T}w_t + o_P(T^{-1/2}),$$

where w_t is mean zero and is a stationary and mixing process. The CLT follows.

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

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

We have in general

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

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

At the true parameter values

$$\frac{\partial L(g_0(u);u)}{\partial g} = \frac{1}{T}\sum_{t=1}^{T} K_h(t/T-u)\frac{1}{g(u)}s_2\left(\zeta_t\right)$$

$$\frac{\partial^2 L(g_0(u);u)}{\partial g^2} \simeq \frac{1}{T}\sum_{t=1}^{T} K_h(t/T-u)\frac{1}{g(u)^2}s_2'\left(\zeta_t\right)\zeta_t.$$

We have by integration by parts

$$\begin{aligned} E\left(s_2'\left(\zeta_t\right)\zeta_t\right) &= \int s_2'\left(\zeta\right)\zeta f(\zeta)d\zeta \\ &= -\int s_2\left(\zeta\right)\zeta f'(\zeta)d\zeta - \int s_2\left(\zeta\right) f(\zeta)d\zeta \\ &= -\int s_2^2\left(\zeta\right) f(\zeta)d\zeta \\ &= -I_2(f). \end{aligned}$$

This guarantees that $E(\partial^2 L(g_0(u);u)/\partial g^2) = -I_2(f)/g(u)^2$. The argument for the case with estimated L is similar to Fan and Chen (1999).

Proof of Theorem 5. We show that

$$\widetilde{\eta} - \eta_0 = -\mathcal{I}_T^*(\eta_0, g_0)^{-1} S_T^*(\eta_0, g_0) + o_P(T^{-1/2}), \tag{4}$$

$$\mathcal{I}_T^*(\eta_0, g_0) = \frac{1}{T}\sum_{t=1}^{T} \ell_t^*(\eta, g_0)\ell_{\eta t}^*(\eta, g_0)^\intercal, \quad S_T^*(\eta_0, g_0) = \frac{1}{T}\sum_{t=1}^{T} \ell_{\eta t}^*(\eta_0, g_0),$$

where $\ell^*_{\eta t}(\eta_0, g_0)$ is an MDS, and the result follows by LLN and CLT for stationary mixing processes. The approximation (4) follows by Tayor series expansions using the smoothness and moment conditions. By construction of the efficient score function, the contribution from $\widehat{g}$ is not present.

Specifically, we show that for any sequence $\eta_T = \eta_0 + T^{-1/2}w$ for $w \in \mathbb{R}^3$

$$S^*_T(\eta_T, \widehat{g}) - S^*_T(\eta_T, g_0) = o_P(T^{-1/2})$$

$$\mathcal{I}^*_T(\widehat{\eta}, \widehat{g}) - \mathcal{I}^*_T(\eta_0, g_0) = o_P(1)$$

and then apply standard arguments from Kreiss (1987) and Linton (1993).

In this argument we use second order expansions, in particular,

$$\begin{aligned}
\widehat{\zeta}_t - \zeta_t &= \frac{\ell_t}{\widehat{g}(t/T)\widehat{\lambda}_t} - \frac{\ell_t}{g(t/T)\lambda_t} \qquad (5)\\
&= -\zeta_t\left(\frac{\widehat{\lambda}_t - \lambda_t}{\lambda_t}\right) + \zeta_t\left(\frac{\widehat{\lambda}_t - \lambda_t}{\lambda_t}\right)^2 - \zeta_t\left(\frac{\widehat{g}(t/T) - g(t/T)}{g(t/T)}\right) + \zeta_t\left(\frac{\widehat{g}(t/T) - g(t/T)}{g(t/T)}\right)^2\\
&+ \zeta_t\left(\frac{\widehat{\lambda}_t - \lambda_t}{\lambda_t}\right)\left(\frac{\widehat{g}(t/T) - g(t/T)}{g(t/T)}\right) + Rem_{t,T},
\end{aligned}$$

where $Rem_{t,T}$ is a remainder term that is of smaller order. We then further replace $\widehat{\lambda}_t - \lambda_t$ by the leading terms of (2). The quadratic terms are all bounded using the uniform rate of convergence of $\widehat{g}(u) - g(u)$ and the root-n consistency of $\widehat{\theta}$. We likewise expand $\widehat{\varphi}$ around its limit and obtain terms of the form

$$\begin{aligned}
\widehat{s}_2(\zeta) - s_2(\zeta) &= \zeta\left(\frac{\partial \log f_{\widehat{\varphi}}}{\partial \zeta}(\zeta) - \frac{\partial \log f_{\varphi}}{\partial \zeta}(\zeta)\right)\\
&= \zeta\frac{\partial}{\partial \varphi}\left(\frac{\partial \log f_{\varphi}}{\partial \zeta}\right)(\zeta)\,(\widehat{\varphi} - \varphi) + \frac{1}{2}\zeta\frac{\partial^2}{\partial \varphi^2}\left(\frac{\partial \log f_{\varphi}}{\partial \zeta}\right)(\zeta)\,(\widehat{\varphi} - \varphi)^2\\
&+ \frac{1}{2}\zeta\frac{\partial^2}{\partial \varphi^2}\left(\frac{\partial \log f_{\overline{\varphi}}}{\partial \zeta}\right)(\zeta)\,(\widehat{\varphi} - \varphi)^2 - \frac{1}{2}\zeta\frac{\partial^2}{\partial \varphi^2}\left(\frac{\partial \log f_{\varphi}}{\partial \zeta}\right)(\zeta)\,(\widehat{\varphi} - \varphi)^2
\end{aligned}$$

for some $\overline{\varphi}$ such that $|\overline{\varphi} - \varphi| \leq |\widehat{\varphi} - \varphi| = O_P(T^{-1/2})$. It follows that when $|\widehat{\varphi} - \varphi| \leq CT^{-1/2}$

$$\begin{aligned}
&\sup_{|\zeta| \leq Q_T}\left|\widehat{s}_2(\zeta) - s_2(\zeta) - \zeta\frac{\partial}{\partial \varphi}\left(\frac{\partial \log f_{\varphi}}{\partial \zeta}\right)(\zeta)\,(\widehat{\varphi} - \varphi) - \frac{1}{2}\zeta\frac{\partial^2}{\partial \varphi^2}\left(\frac{\partial \log f_{\varphi}}{\partial \zeta}\right)(\zeta)\,(\widehat{\varphi} - \varphi)^2\right|\\
&= \frac{1}{2}Q_T \times CT^{-1}.
\end{aligned}$$

The leading terms fit into sample averages and can be analyzed by laws of large numbers. Regarding the remainder term, we have by the Bonferroni and Markov inequalities

$$\Pr\left(\max_{1 \leq t \leq T} \zeta_t R(\zeta_t) > Q_T\right) \leq T \Pr\left(\zeta_t R(\zeta_t) > Q_T\right) \leq T\frac{E(\left(\zeta_t R(\zeta_t)\right)^{\kappa})}{Q_T^{\kappa}} = o(1)$$

provided $Q_T = T^{1/\kappa}/\log T$. Therefore, with $\kappa = 4$, $Q_T T^{-1/2} \to 0$ and the remainder term is $o_P(T^{-1/2})$.

Proof of Theorem 6. We show that

$$\widetilde{\widetilde{\theta}} - \theta_0 = -\mathcal{I}_T^{**}(\theta_0, f_0, g_0)^{-1} S_T^{**}(\theta_0, f_0, g_0) + o_P(T^{-1/2})$$

$$\mathcal{I}_T^{**}(\theta_0, f_0, g_0) = \frac{1}{T}\sum_{t=1}^{T} \ell_{\theta t}^{**}(\theta_0, f, g)\ell_{\theta t}^{**}(\theta_0, f, g)^{\intercal}, \quad S_T^{**}(\theta_0, f_0, g_0) = \frac{1}{T}\sum_{t=1}^{T} \ell_{\theta t}^{**}(\theta_0, f, g).$$

The arguments are lengthy and repeated in many places in the literature. Furthermore, they often use additional devices like sample splitting and discretization. We first discuss the trimming issue. Since the density f has unbounded support on the right side, it is necessary to trim out the contributions where f is small; this argument is presented in Linton and Xiao (2007) using "smooth trimming". Specifically, let $\tau(\cdot)$ be a density function that has support $[0, 1]$, $\tau(0) = \tau(1) = 0$, and let

$$\tau_b(x) = \frac{1}{b}\tau\left(\frac{x}{b} - 1\right),$$

where b is the trimming parameter; then $\tau_b(x)$ has support on $[b, 2b]$. Letting $\Upsilon_b(x) = \int_0^x \tau_b(z)dz$, we have

$$\Upsilon_b(x) = \begin{cases} 0, & x < b \\ \int_{-\infty}^{x} \tau_b(z)dz, & b \le x \le 2b \\ 1, & x > 2b. \end{cases}$$

For example, consider the following Beta density $\tau(z) = B(a+1)^{-1} z^a (1-z)^a, \quad 0 \le z \le 1$, for some positive integer a, where $B(a)$ is the beta function defined by $B(a) = \Gamma(a)^2/\Gamma(2a)$, and $\Gamma(a)$ is the Euler gamma function. Then, it can be verified that the function $\Upsilon_b(x)$ is $(a+1)-$times continuously differentiable on $[0, 1]$. This property allows us to use standard Taylor series arguments, whereas indicator function trimming would preclude this. We will assume that $a \ge 3$. with some function Υ_b. Then let $\widehat{1}_t = \Upsilon_b(\widehat{f}(\widehat{\zeta}_t))$, and define

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

for any $\theta \in \Theta$.

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. Consider the log likelihood

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

$$\lambda_t(\theta, \delta) = 1 - \beta - \gamma + \beta\lambda_{t-1}(\theta, \delta) + \gamma \frac{\ell_{t-1}}{g_\delta((t-1)/T)}, \tag{6}$$

$$\zeta_t(\theta, \delta) = \frac{\ell_t}{\lambda_t(\theta, \delta) g_\delta(t/T)}. \tag{7}$$

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,

$$\frac{\partial \lambda_t(\theta, \delta)}{\partial \beta} = \beta \frac{\partial \lambda_{t-1}(\theta, \delta)}{\partial \beta} + \lambda_t(\theta, \delta) - 1$$

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

and so

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

Likewise,

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

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

and so

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

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,

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

$$\begin{aligned}\frac{\partial \lambda_t(\theta,\delta)}{\partial \delta} &= -\gamma \lambda_{t-1}\zeta_{t-1}\frac{\partial \log g_\delta((t-1)/T)}{\partial \delta} - \beta\gamma \lambda_{t-2}\zeta_{t-2}\frac{\partial \log g_\delta((t-2)/T)}{\partial \delta} \\ &\quad - \beta^2\gamma \lambda_{t-3}\zeta_{t-3}\frac{\partial \log g_\delta((t-3)/T)}{\partial \delta} - \dots\end{aligned}$$

Then we expand

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

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

$$\begin{aligned}\sum_{j=1}^T \psi_j g\left(\frac{t-j}{T}\right) &= g\left(\frac{t}{T}\right)\sum_{j=1}^T \psi_j - g'\left(\frac{t}{T}\right)\frac{1}{T}\sum_{j=1}^T \psi_j j - \frac{1}{2T}\sum_{j=1}^T \psi_j j^2 g''\left(\frac{s^*(t,j)}{T}\right) \\ &\simeq g\left(\frac{t}{T}\right)\sum_{j=1}^T \psi_j.\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{\partial L}{\partial \delta} &= \sum_{t=1}^T \frac{f'}{f}(\zeta_t)\frac{\partial \zeta_t}{\partial \delta} - \frac{\partial \log \lambda_t}{\partial \delta} - \frac{\partial \log g_\delta(t/T)}{\partial \delta} \\ &= \sum_{t=1}^T s_2(\zeta_t)\left(\frac{\partial \log \lambda_t}{\partial \delta} + \frac{\partial \log g_\delta(t/T)}{\partial \delta}\right) \qquad (8) \\ &= \sum_{t=1}^T s_2(\zeta_t)\frac{\partial \log g_\delta(t/T)}{\partial \delta}\left(1 - \gamma\frac{(1-\beta L)^{-1}\lambda_{t-1}\zeta_{t-1}}{\lambda_t}\right) \\ &= \frac{1-\beta-\gamma}{1-\beta}\sum_{t=1}^T s_2(\zeta_t)\frac{1}{\lambda_t}\frac{\partial \log g_\delta(t/T)}{\partial \delta}, \qquad (9)\end{aligned}$$

since

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

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

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

In conclusion, the tangent space for g consists of functions of the form

$$\mathbb{T}_g = \left\{ \sum_{t=1}^{T} s_2(\zeta_t)\frac{1}{\lambda_t}h(t/T) : h \in L_2[0,1] \right\}. \tag{10}$$

That is, the score w.r.t. g is of the form

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

for some function $h(.)$ and the information is of the form

$$\frac{1}{T}\sum_{t=1}^{T} I_2(f)E\left(\frac{1}{\lambda_t}\right)h(t/T)^2 \sim I_2(f)E\left(\frac{1}{\lambda_t}\right)\int h(u)^2du$$

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

$$\begin{aligned} L_\theta^* &= \sum_{t=1}^{T} s_2(\zeta_t)\left(\frac{\partial \log\lambda_t}{\partial\theta} - \frac{E\left[\frac{\partial\log\lambda_t}{\partial\theta}\frac{1}{\lambda_t}\right]}{E\left(\frac{1}{\lambda_t^2}\right)}\frac{1}{\lambda_t}\right) \\ &= \sum_{t=1}^{T} s_2(\zeta_t)\frac{1}{\lambda_t}\left(\frac{\partial\lambda_t}{\partial\theta} - \frac{E\left[\frac{\partial\lambda_t}{\partial\theta}\frac{1}{\lambda_t^2}\right]}{E\left(\frac{1}{\lambda_t^2}\right)}\right). \end{aligned} \tag{11}$$

Note the term involving $h(t/T)$ drops out as this is arbitrary. This can be verified, as for any element of $\mathbb{T}_g$ (indexed by $h(.)$) we have

$$\begin{aligned} &\sum_{t=1}^{T} E\left[s_2(\zeta_t)\frac{1}{\lambda_t}\left(\frac{\partial\lambda_t}{\partial\theta} - \frac{E\left[\frac{\partial\lambda_t}{\partial\theta}\frac{1}{\lambda_t^2}\right]}{E\left(\frac{1}{\lambda_t^2}\right)}\right) \times s_2(\zeta_t)\frac{1}{\lambda_t}h(t/T)\right] \\ &= \sum_{t=1}^{T} E\left[s_2^2(\zeta_t)\right] E\left[\frac{1}{\lambda_t^2}\left(\frac{\partial\lambda_t}{\partial\theta} - \frac{E\left[\frac{\partial\lambda_t}{\partial\theta}\frac{1}{\lambda_t^2}\right]}{E\left(\frac{1}{\lambda_t^2}\right)}\right)\right] h(t/T) \\ &= I_2(f)\sum_{t=1}^{T}\left[E\left(\frac{1}{\lambda_t^2}\frac{\partial\lambda_t}{\partial\theta}\right) - \frac{E\left[\frac{\partial\lambda_t}{\partial\theta}\frac{1}{\lambda_t^2}\right]}{E\left(\frac{1}{\lambda_t^2}\right)}E\left(\frac{1}{\lambda_t^2}\right)\right] h(t/T) \\ &= 0. \end{aligned}$$

C.2 Parametric f

Now suppose that $f = f_\varphi$, where φ is unknown. The full parametric likelihood is now

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

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

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

These score functions satisfy the two moment conditions:

$$\int \frac{\partial \log f_\varphi(\zeta_t)}{\partial \varphi} f_\varphi(\zeta)\, d\zeta = \int \frac{\partial f_\varphi(\zeta)}{\partial \varphi} d\zeta = \frac{\partial}{\partial \varphi} \int f_\varphi(\zeta)\, d\zeta = \frac{\partial}{\partial \varphi} 1 = 0.$$

$$\int \zeta \frac{\partial \log f_\varphi(\zeta_t)}{\partial \varphi} f_\varphi(\zeta)\, d\zeta = \int \zeta \frac{\partial f_\varphi(\zeta)}{\partial \varphi} d\zeta = \frac{\partial}{\partial \varphi} \int \zeta f_\varphi(\zeta)\, d\zeta = \frac{\partial}{\partial \varphi} 1 = 0.$$

Furthermore,

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

We have

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

We have

$$\frac{\partial}{\partial \varphi} \int \zeta \frac{f'_\varphi}{f_\varphi}(\zeta) f_\varphi(\zeta)\, d\zeta = \frac{\partial}{\partial \varphi} \int \zeta f'_\varphi(\zeta) d\zeta = 0,$$

and by the Chain rule

$$\int \zeta \frac{\partial}{\partial \varphi}\left(\frac{f'_\varphi}{f_\varphi}(\zeta) f_\varphi(\zeta)\right) d\zeta = \int \zeta \frac{\partial}{\partial \varphi}\left(\frac{f'_\varphi}{f_\varphi}(\zeta)\right) f_\varphi(\zeta)\, d\zeta + \int \zeta \frac{f'_\varphi}{f_\varphi}(\zeta) \frac{\partial f_\varphi(\zeta)}{\partial \varphi} d\zeta$$

so that

$$\int \zeta \frac{f'_\varphi}{f_\varphi}(\zeta) \frac{\partial f_\varphi(\zeta)}{\partial \varphi} d\zeta = -E\left(\zeta \frac{\partial}{\partial \varphi}\left(\frac{f'_\varphi}{f_\varphi}(\zeta)\right)\right).$$

Therefore,

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

for any parameterization of g. We conjecture that the efficient score function for φ in the presence of unknown g is

$$L_{\varphi}^{*} = \sum_{t=1}^{T}\left(\frac{\partial \log f_{\varphi}(\zeta_t)}{\partial \varphi} - \frac{E\left(\frac{\partial \log f_{\varphi}(\zeta_t)}{\partial \varphi} s_2(\zeta_t)\right)}{I_2(f)} \frac{E\left(\frac{1}{\lambda_t}\right)}{E\left(\frac{1}{\lambda_t^2}\right)} s_2(\zeta_t)\frac{1}{\lambda_t}\right). \tag{12}$$

This can be verified since

$$\begin{aligned}
&\sum_{t=1}^{T} E\left(\left(\frac{\partial \log f_{\varphi}(\zeta_t)}{\partial \varphi} - \frac{E\left(\frac{\partial \log f_{\varphi}(\zeta_t)}{\partial \varphi} s_2(\zeta_t)\right)}{I_2(f)} \frac{E\left(\frac{1}{\lambda_t}\right)}{E\left(\frac{1}{\lambda_t^2}\right)} s_2(\zeta_t)\frac{1}{\lambda_t}\right) s_2(\zeta_t)\frac{1}{\lambda_t} h(t/T)\right) \\
&= \sum_{t=1}^{T} E\left(\frac{\partial \log f_{\varphi}(\zeta_t)}{\partial \varphi} s_2(\zeta_t)\frac{1}{\lambda_t} h(t/T)\right) - E\left(\frac{E\left(\frac{\partial \log f_{\varphi}(\zeta_t)}{\partial \varphi} s_2(\zeta_t)\right)}{I_2(f)} \frac{E\left(\frac{1}{\lambda_t}\right)}{E\left(\frac{1}{\lambda_t^2}\right)} s_2(\zeta_t)\frac{1}{\lambda_t} s_2(\zeta_t)\frac{1}{\lambda_t} h(t/T)\right) \\
&= \sum_{t=1}^{T} E\left(\frac{\partial \log f_{\varphi}(\zeta_t)}{\partial \varphi} s_2(\zeta_t)\right) E\left(\frac{1}{\lambda_t}\right) h(t/T) - \sum_{t=1}^{T} E\left(E\left(\frac{\partial \log f_{\varphi}(\zeta_t)}{\partial \varphi} s_2(\zeta_t)\right) E\left(\frac{1}{\lambda_t}\right)\right) h(t/T) \\
&= 0
\end{aligned}$$

for any h.

C.3 Unknown f

We next consider the semiparametric case where f is of unknown form but has unit mean. According to Drost and Werker (2004), the tangent space for f consists of functions τ that satisfy

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

Recall that the tangent space for g consists of functions of the form

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

and these two spaces are not orthogonal. We must project

$$L_{\theta} = \sum_{t=1}^{T} s_2(\zeta_t)\frac{\partial \log \lambda_t}{\partial \theta}$$

orthogonally to their union. Formally, one may write

$$L_\theta^{**} = L_\theta - ACE\left(L_\theta \left| \mathbb{T}_g + \mathbb{T}_f \right.\right),$$

where $ACE(.|.)$ is the alternating conditional expectation operator, see Bickel et al. (1993) (Proposition 1).

According to Drost and Werker (2004) (Example 3, iid errors) the projection of L_θ orthogonal to the tangent space $\mathbb{T}_f$ is

$$\sum_{t=1}^{T} \frac{\zeta_t - 1}{\sigma_\zeta^2} E\left(\frac{\partial \log \lambda_t}{\partial \theta}\right) + \sum_{t=1}^{T} s_2(\zeta_t)\left(\frac{\partial \log \lambda_t}{\partial \theta} - E\left(\frac{\partial \log \lambda_t}{\partial \theta}\right)\right).$$

This score function is not orthogonal to $\mathbb{T}_g$ so it is not a candidate here.

In the sequel we make use of the fact that

$$E\left((\zeta_t - 1)\, s_2(\zeta_t)\right) = E\left(\zeta_t s_2(\zeta_t)\right) = 1$$

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

by integration by parts.

We claim that the projection of L_θ onto $\mathbb{T}_f + \mathbb{T}_g$ is of the form

$$\sum_{t=1}^{T}\left(-\frac{\zeta_t - 1}{\sigma_\zeta^2} + s_2(\zeta_t)\right) a + \sum_{t=1}^{T} s_2(\zeta_t) b \frac{1}{\lambda_t} \equiv T_f^* + T_g^*,$$

for some a, b since the first component T_f^* is in $\mathbb{T}_f$ and the second component T_g^* is in $\mathbb{T}_g$. It follows that the efficient score function is of the form

$$L_\theta^{**} = L_\theta - (T_f^* + T_g^*) = \sum_{t=1}^{T} \frac{\zeta_t - 1}{\sigma_\zeta^2} a + \sum_{t=1}^{T} s_2(\zeta_t)\left(\frac{\partial \log \lambda_t}{\partial \theta} - a - b\frac{1}{\lambda_t}\right). \tag{13}$$

This is orthogonal to both $\mathbb{T}_g$ and $\mathbb{T}_f$ if and only if:

$$E\left(\frac{\partial \log \lambda_t}{\partial \theta} - a - b\frac{1}{\lambda_t}\right) = 0 \tag{14}$$

$$E\left(\frac{1}{\sigma_\zeta^2} a \frac{1}{\lambda_t} + I_2(f)\left(\frac{1}{\lambda_t}\frac{\partial \log \lambda_t}{\partial \theta} - a\frac{1}{\lambda_t} - b\frac{1}{\lambda_t^2}\right)\right) = 0. \tag{15}$$

The first condition arises because we need the second term in L_θ^{**} to be orthogonal to $\mathbb{T}_f$ (the first term is automatically so), while the second condition arises because we need L_θ^{**} to be orthogonal to $\mathbb{T}_g$. We rewrite the second condition as

$$E\left(\left(\frac{1}{\lambda_t}\left(\frac{\partial \log \lambda_t}{\partial \theta} - a\kappa - b\frac{1}{\lambda_t}\right)\right)\right) = 0, \tag{16}$$

where $\kappa = 1 - 1/I_2(f)\sigma_\zeta^2$. Using (14) we must have

$$a = E\left(\frac{\partial \log \lambda_t}{\partial \theta}\right) - bE\left(\frac{1}{\lambda_t}\right).$$

We then substitute into (16) to obtain

$$E\left(\frac{1}{\lambda_t}\frac{\partial \log \lambda_t}{\partial \theta}\right) - \kappa E\left(\frac{\partial \log \lambda_t}{\partial \theta}\right)E\left(\frac{1}{\lambda_t}\right) = b\left(E\left(\frac{1}{\lambda_t^2}\right) - \kappa E^2\left(\frac{1}{\lambda_t}\right)\right)$$

or

$$b = \frac{E\left(\frac{1}{\lambda_t}\frac{\partial \log \lambda_t}{\partial \theta}\right) - \kappa E\left(\frac{\partial \log \lambda_t}{\partial \theta}\right)E\left(\frac{1}{\lambda_t}\right)}{E\left(\frac{1}{\lambda_t^2}\right) - \kappa E^2\left(\frac{1}{\lambda_t}\right)}.$$

In conclusion, for these a, b the efficient score in (13) satisfies the orthogonality condition.

Note that under only the MDS assumption, Drost and Werker (2004) (Example 3) the projection of L_θ orthogonal to the tangent space $\mathbb{T}_f$ is

$$\sum_{t=1}^{T} \frac{\zeta_t - 1}{\operatorname{var}\left(\zeta_t | \mathcal{F}_{t-1}\right)} \frac{\partial \log \lambda_t}{\partial \theta}.$$

This score function is not orthogonal to $\mathbb{T}_g$. The projection orthogonal to $\mathbb{T}_g$ is

$$\sum_{t=1}^{T} \frac{\zeta_t - 1}{\operatorname{var}\left(\zeta_t | \mathcal{F}_{t-1}\right)} \left(\frac{\partial \log \lambda_t}{\partial \theta} - b\frac{1}{\lambda_t}\right),$$

where b is the slope of the best linear no intercept predictor,

$$b = \frac{E\left(\frac{\partial \log \lambda_t}{\partial \theta}\frac{1}{\lambda_t}\right)}{E\left(\frac{1}{\lambda_t^2}\right)}.$$

This is the efficient score function in the model where only the conditional moment restriction is made.

D Risk premium regressions

We consider the population regression model

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

where in practice we replace $m(.)$ by $\widehat{m}(.)$ 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

$$y_t = \beta^{\intercal} x_t(\theta_0) + \varepsilon_t.$$

By a Taylor expansion we have

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

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

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

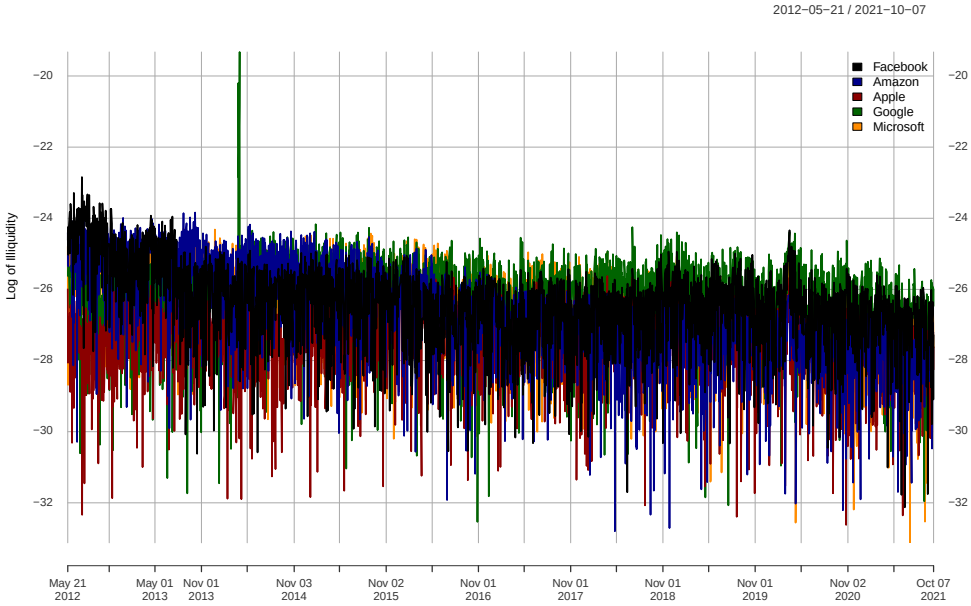
Then we regress $\widehat{y}_t$ on x_t^* and use the linear regression standard errors.

E Other tables and figures

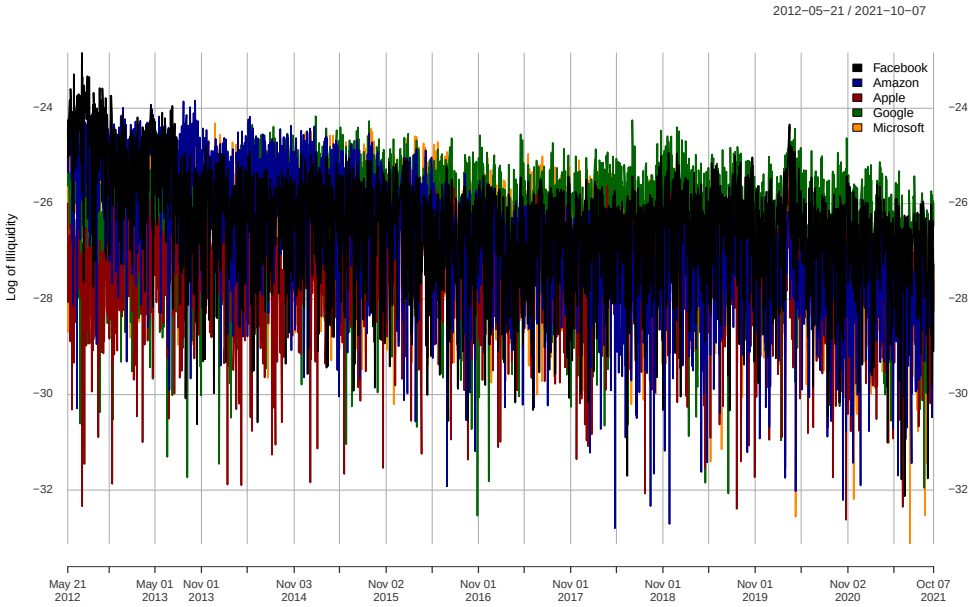
E.1 Amihud illiquidity

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 a 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.

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



(a) Original data.



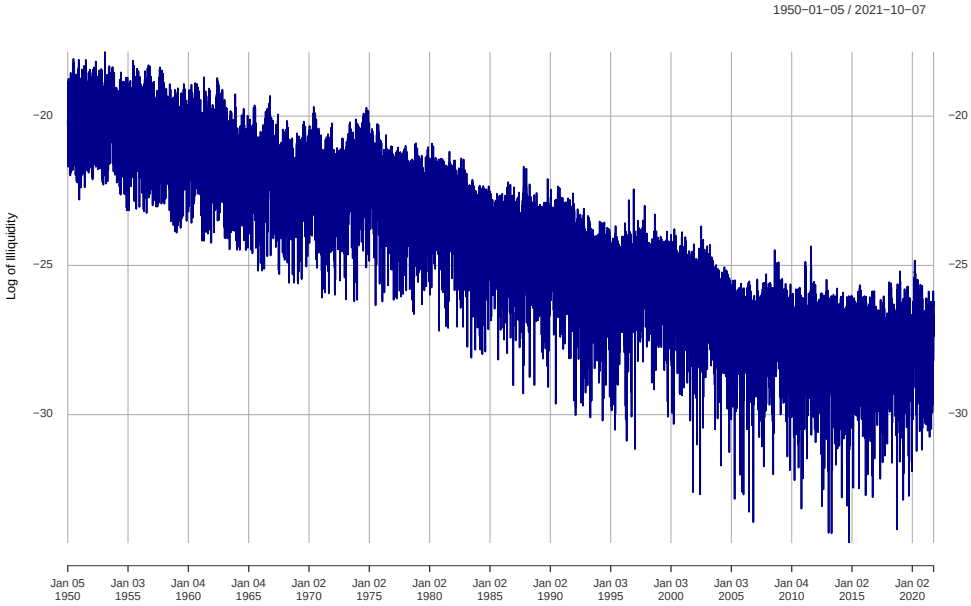
(b) Data of Google stock processed.

Figure 1: Fab 5 daily log illiquidity – $\log \ell_t$.

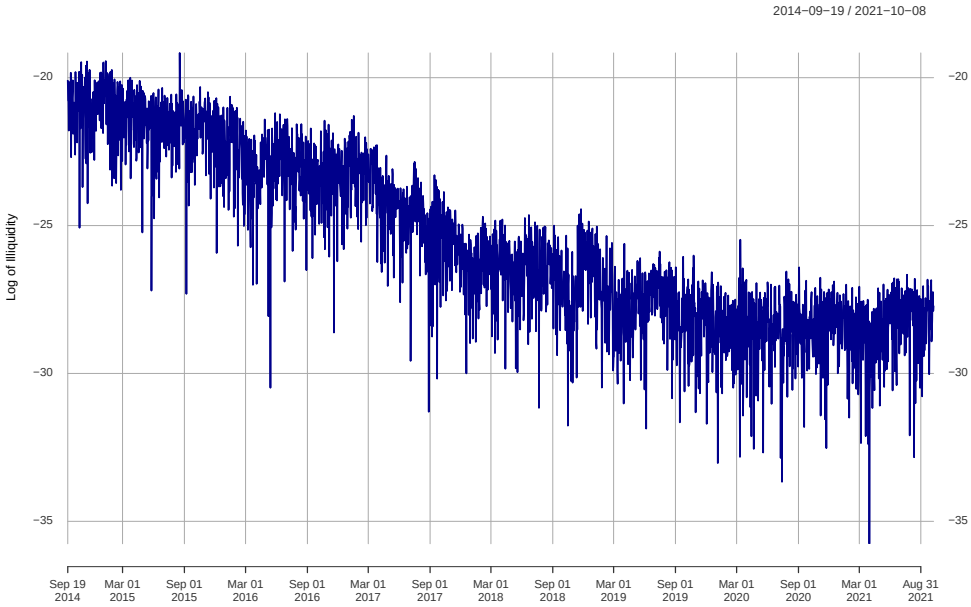
Table 1: Estimated parameters of an AR(5) with trend.

	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&P 500 index.



(b) Bitcoin asset.

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

E.2 Estimation of long-run trend function

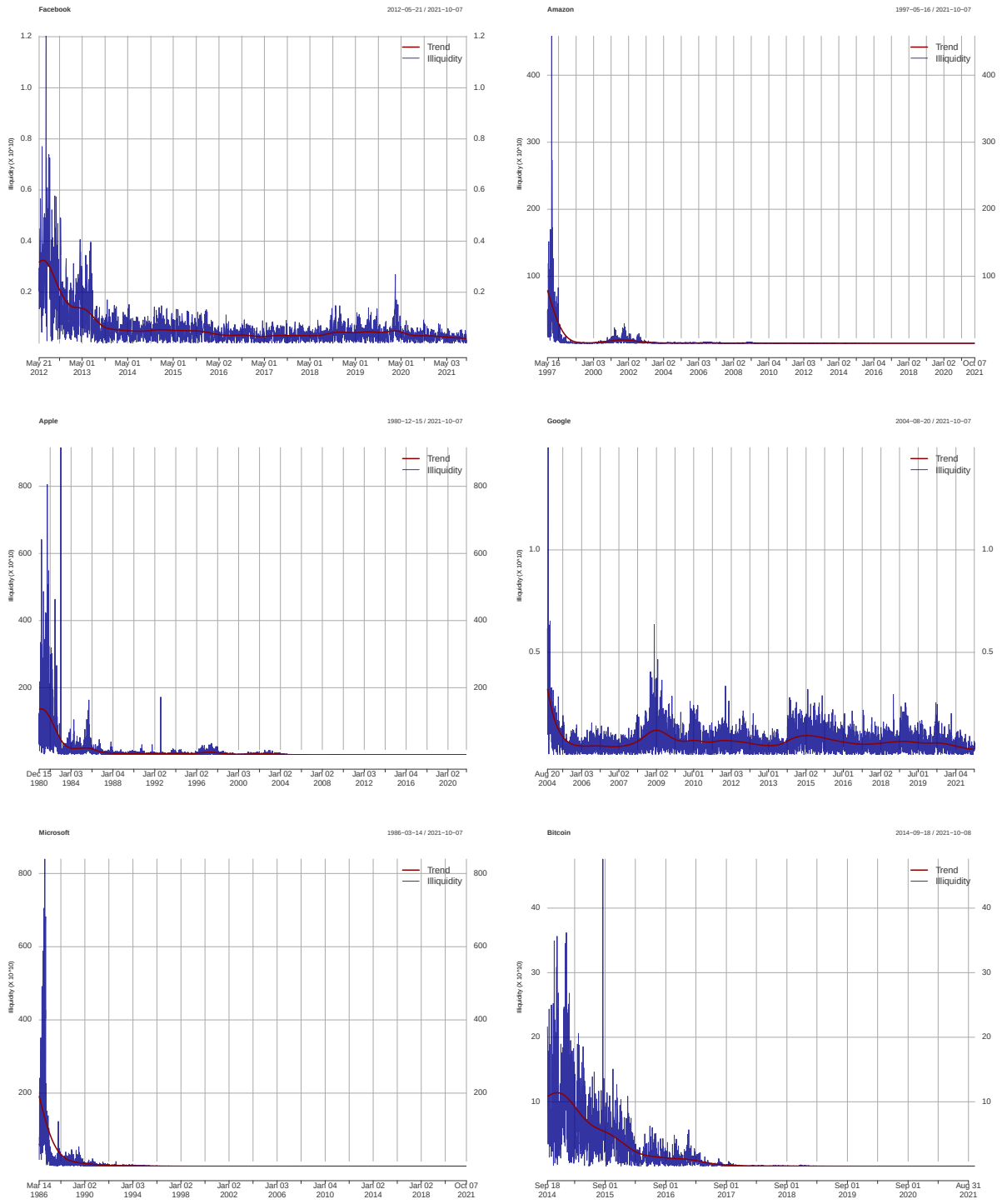


Figure 3: Fab 5 and Bitcoin illiquidity series and trend functions ($\times 10^{10}$).

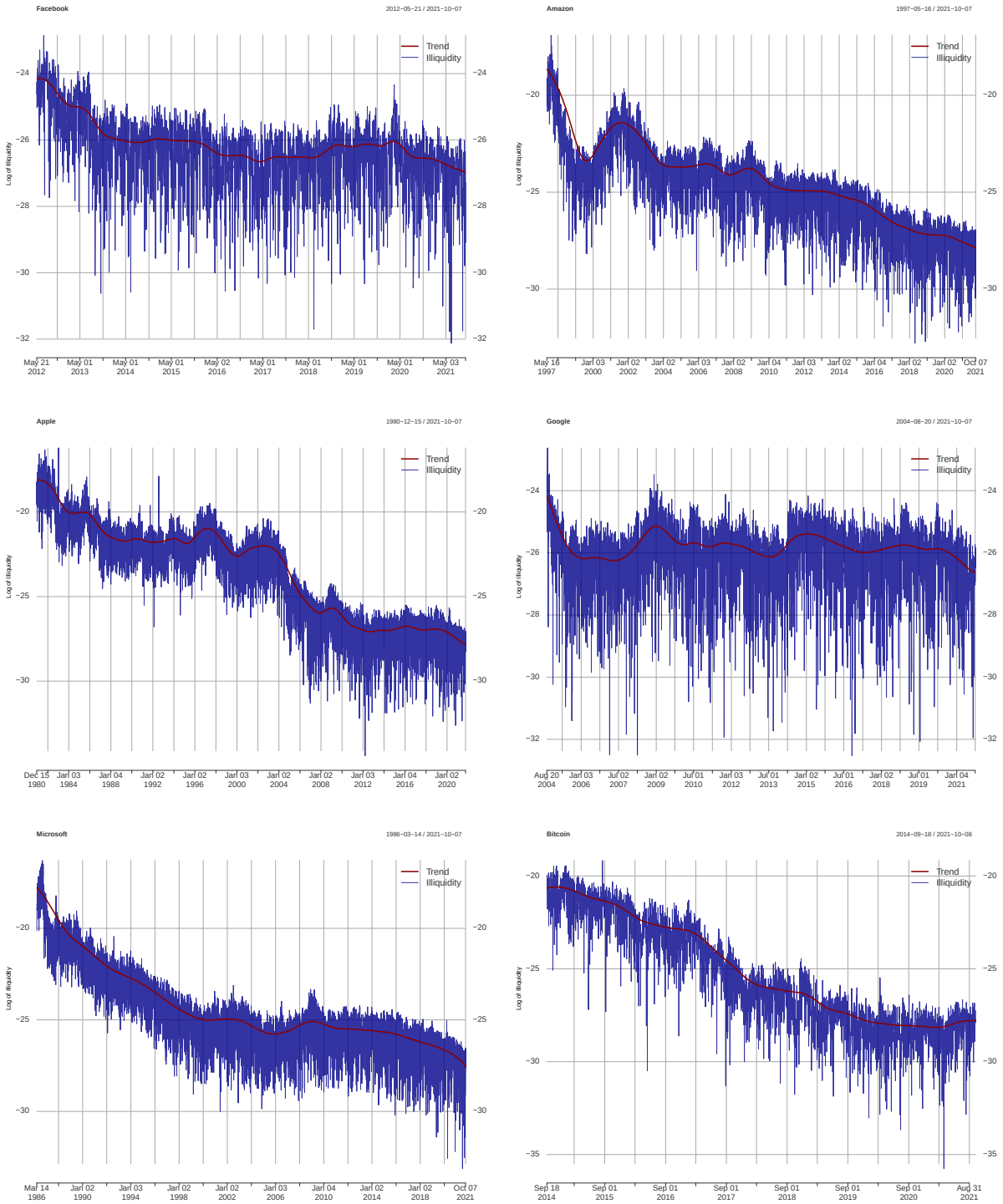


Figure 4: Fab 5 and Bitcoin log illiquidity series and trend functions.

E.3 Estimation based on conditional moment restrictions

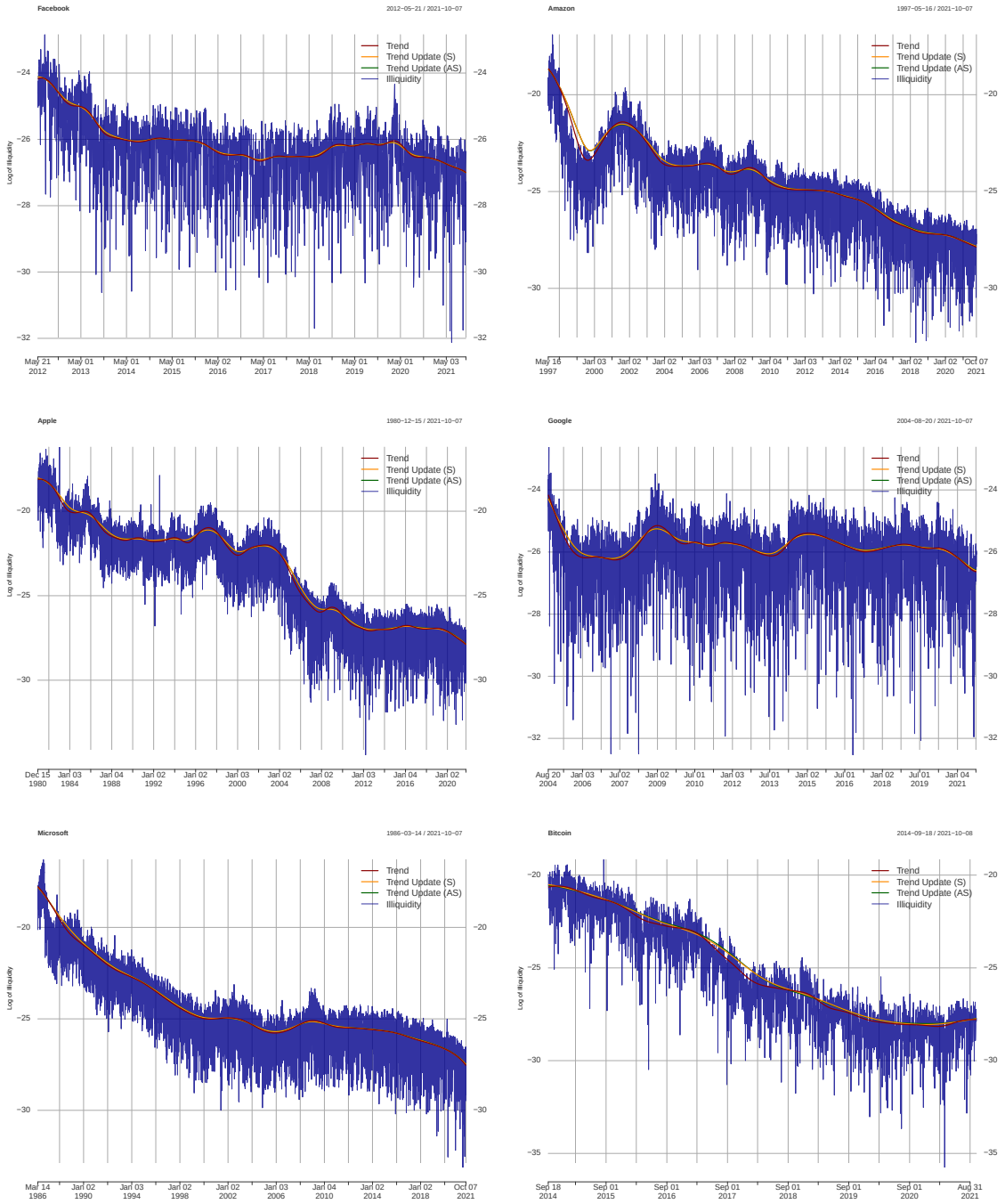


Figure 5: Fab 5 and Bitcoin log illiquidity and updated trend function based on the GMM estimator of λ_t parameters. The red curve corresponds to the initial estimate of the trend function and the yellow and green curves correspond to the updated estimates in, respectively, the symmetric and asymmetric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

E.4 Estimation: i.i.d. error term with parametric density

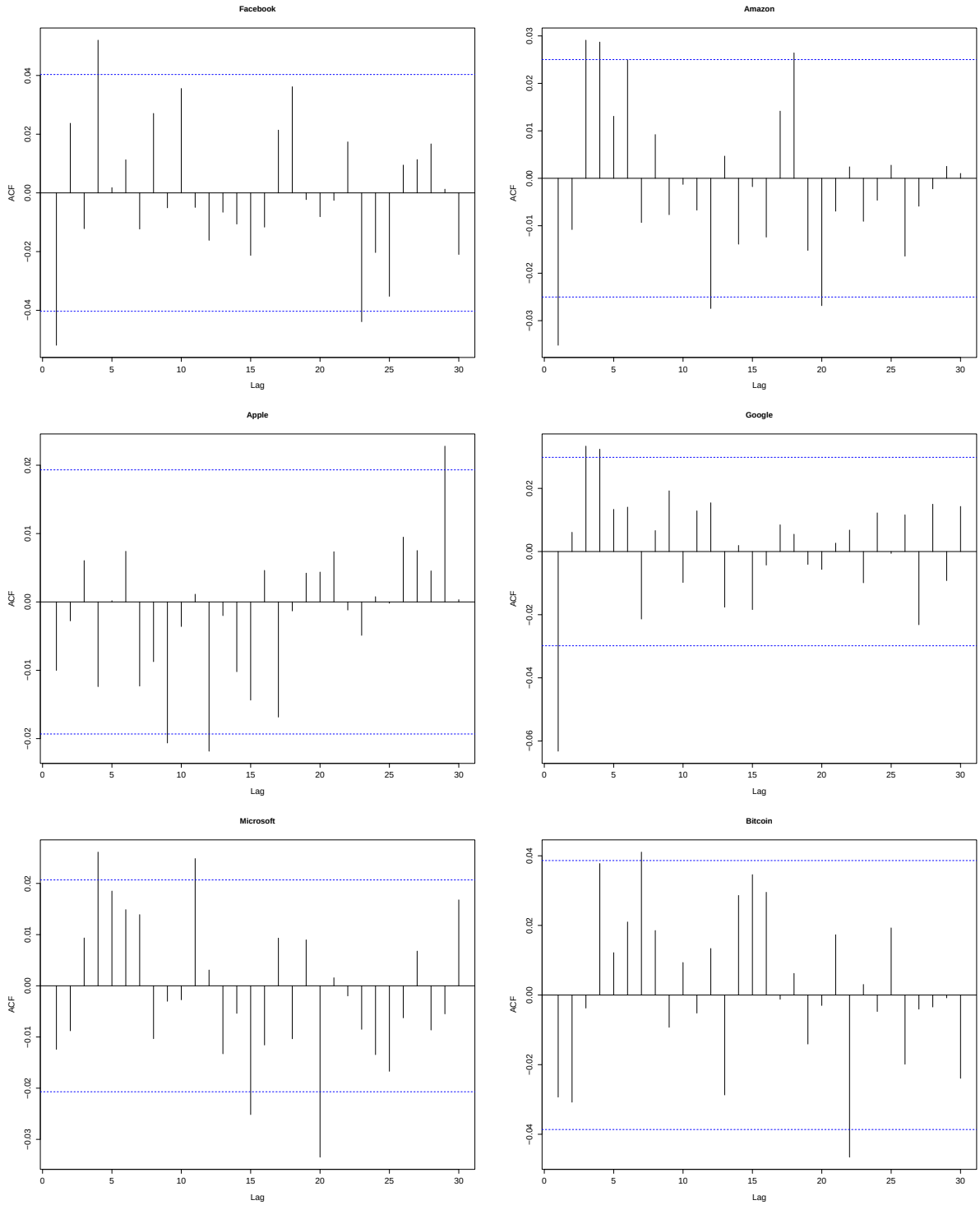


Figure 6: ACF of ζ_t under the symmetric specification for λ_t.

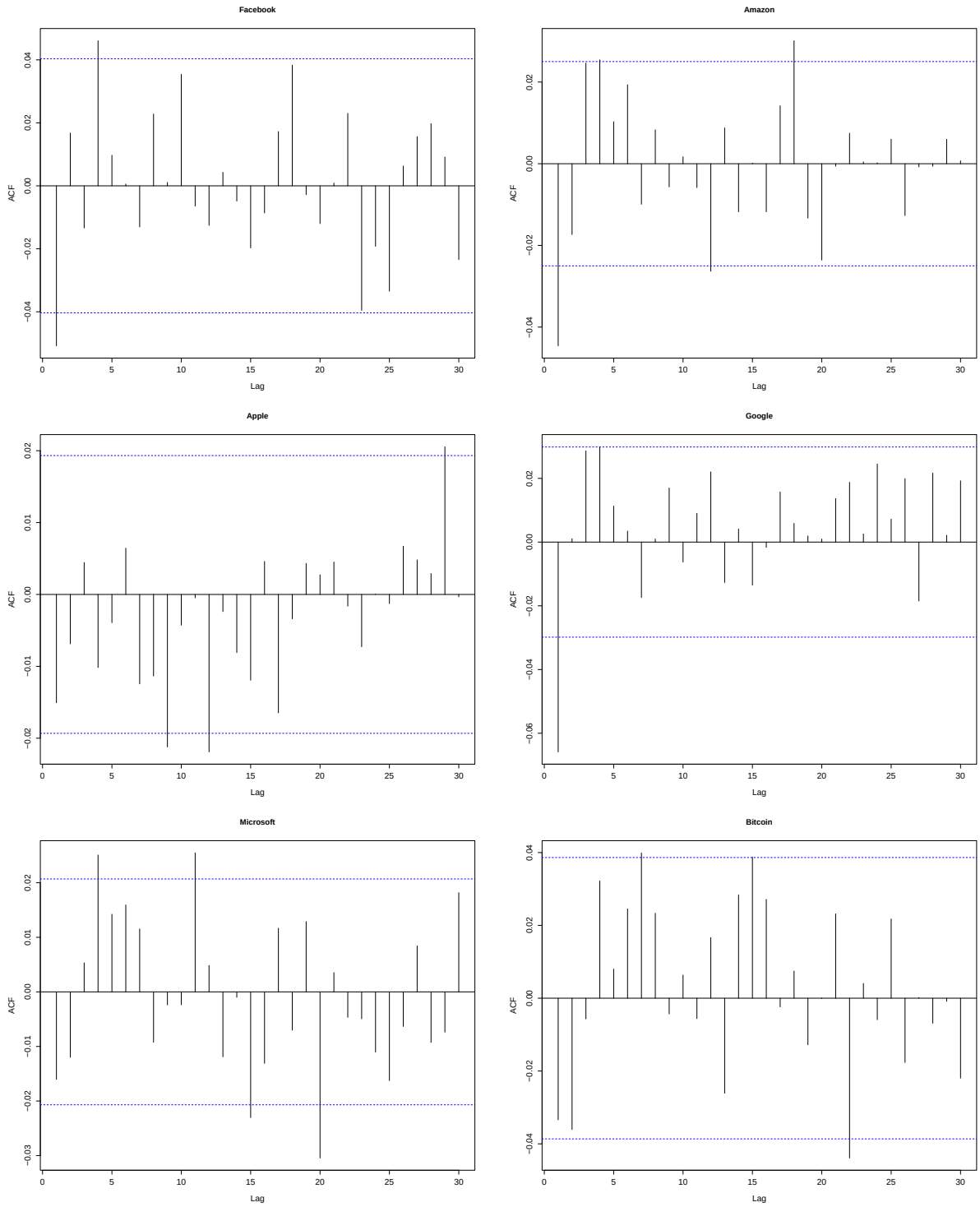


Figure 7: ACF of ζ_t under the asymmetric specification for λ_t.

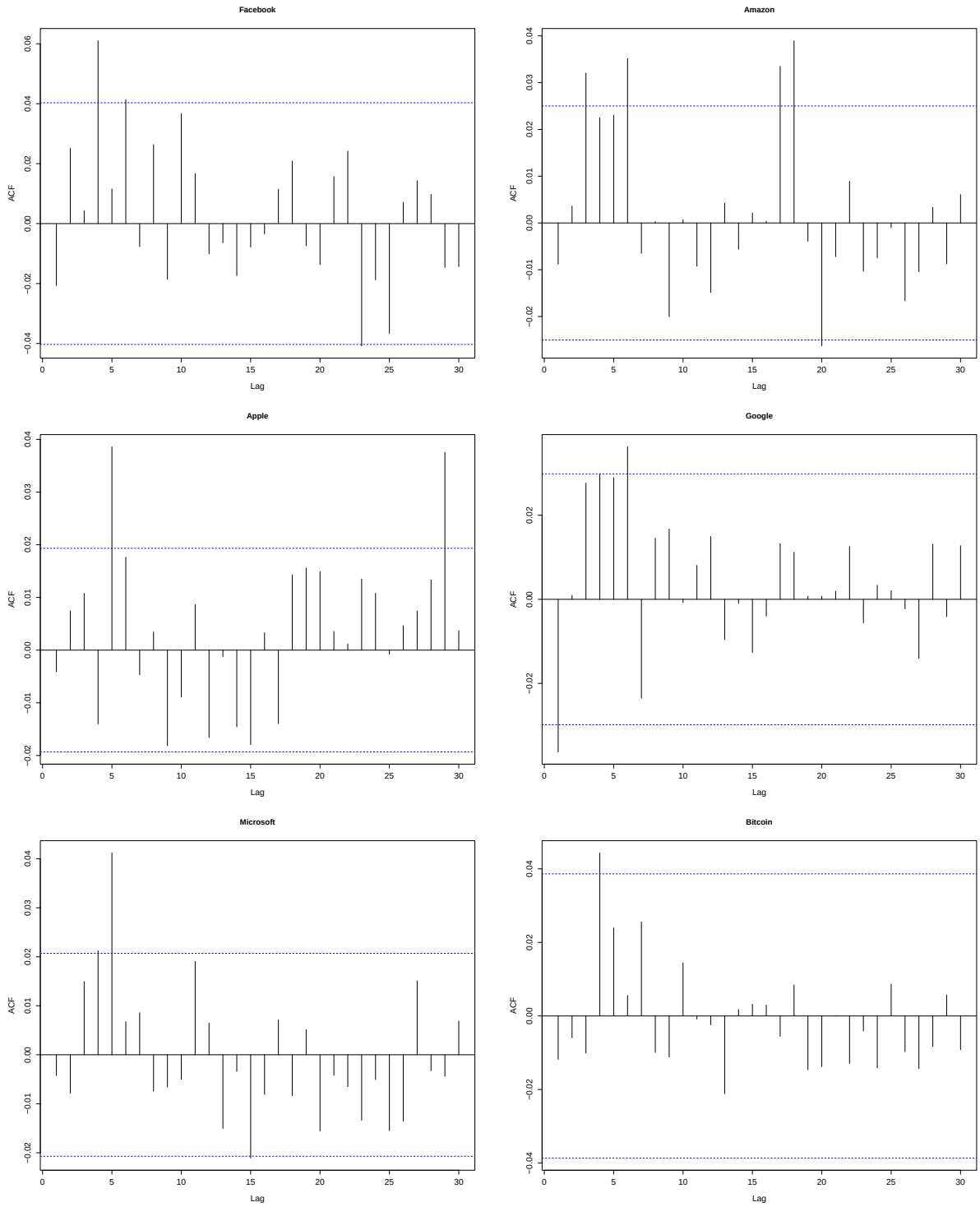


Figure 8: ACF of ζ_t^2 under the symmetric specification for λ_t.

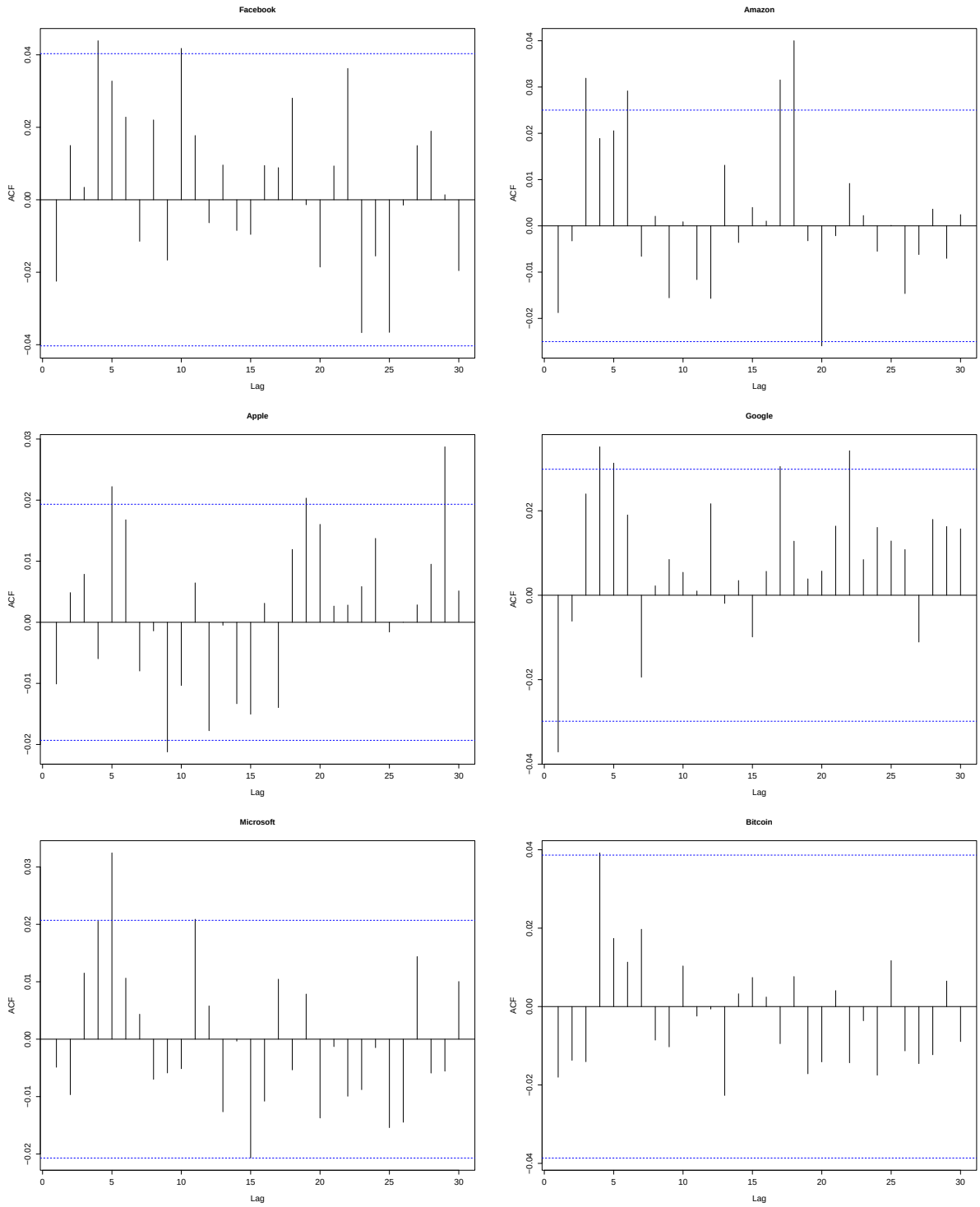


Figure 9: ACF of ζ_t^2 under the asymmetric specification for λ_t.

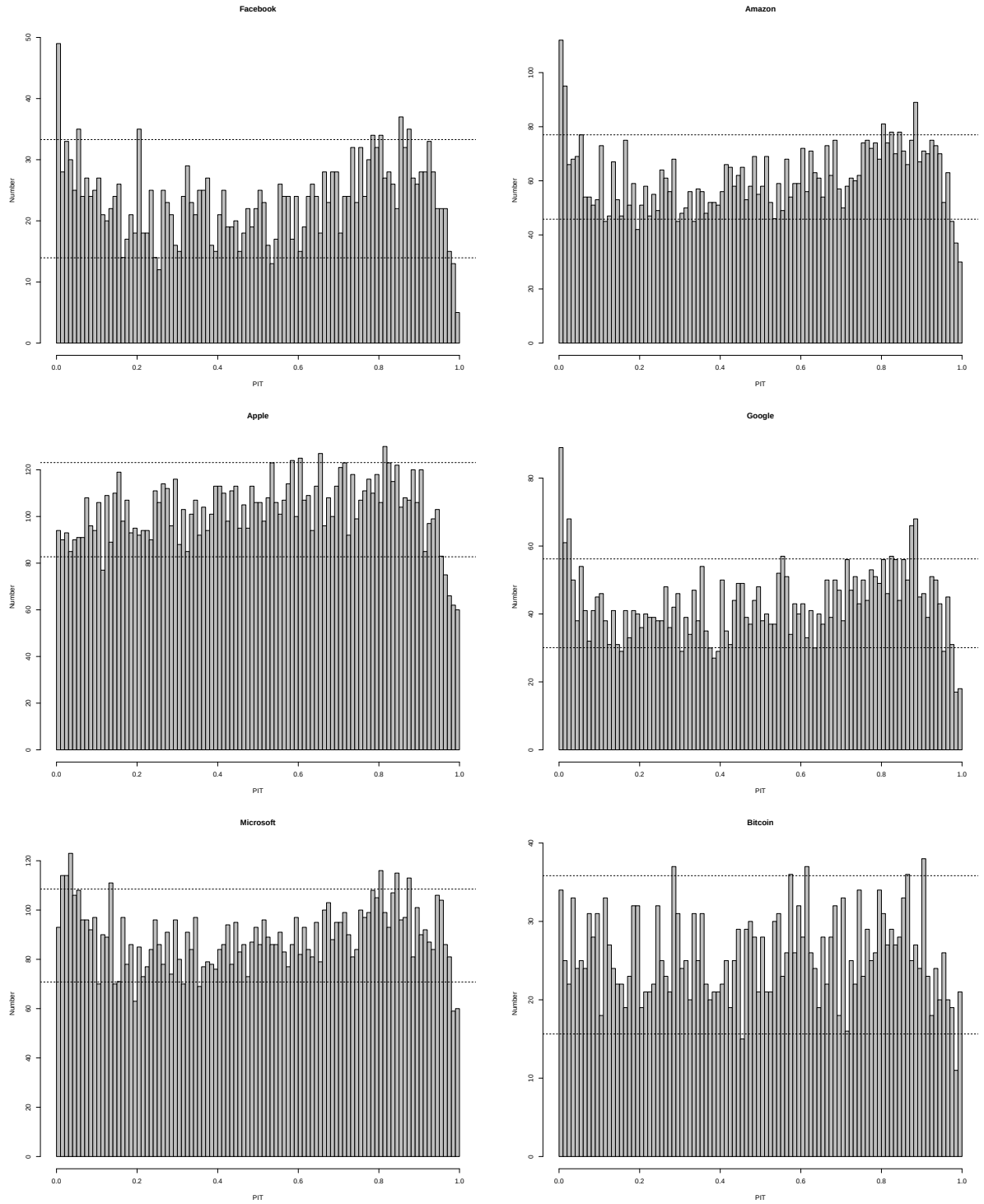


Figure 10: Probability integral transform (PIT) of ζ_t under the symmetric specification for λ_t.

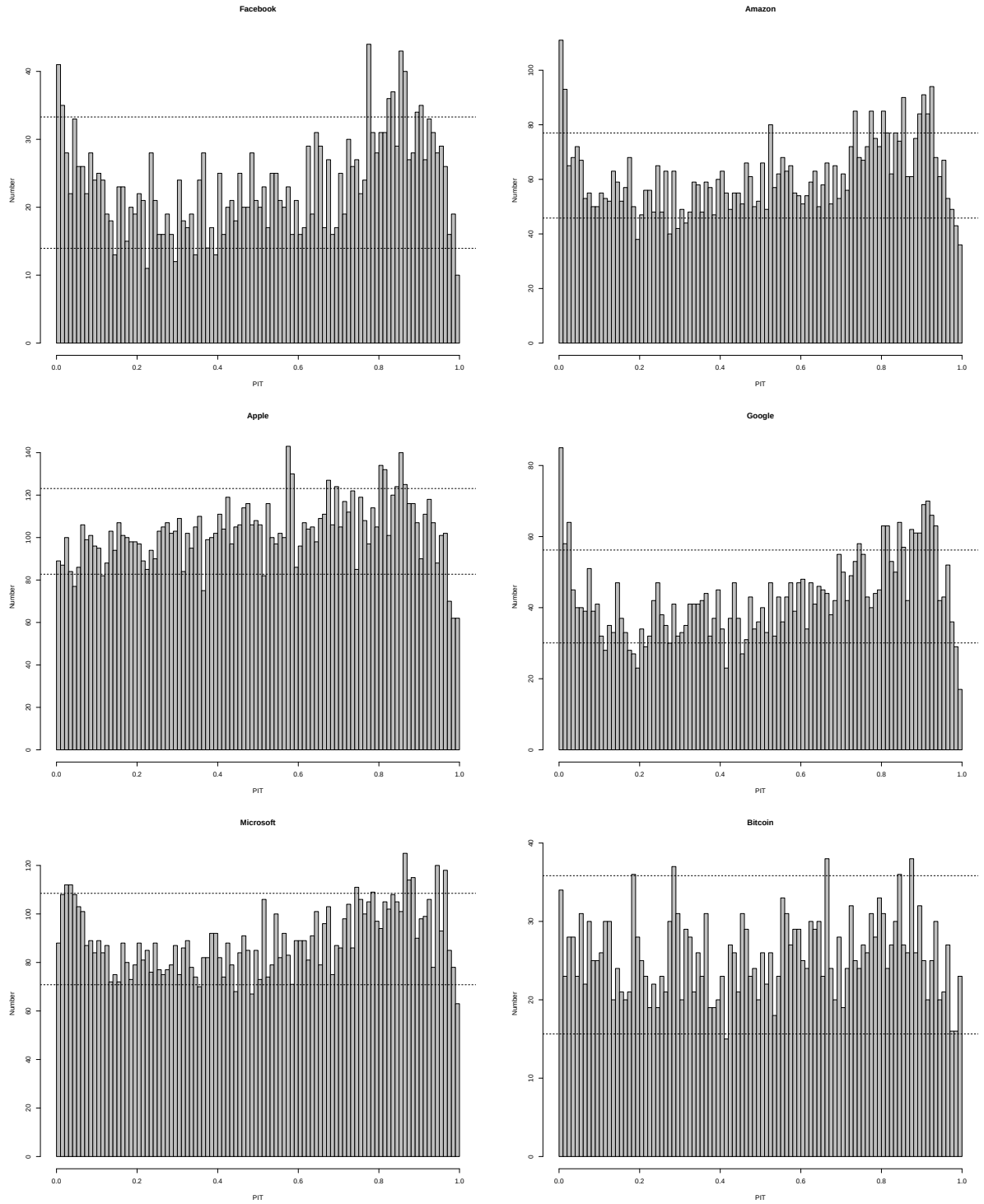


Figure 11: Probability integral transform (PIT) of ζ_t under the asymmetric specification for λ_t.

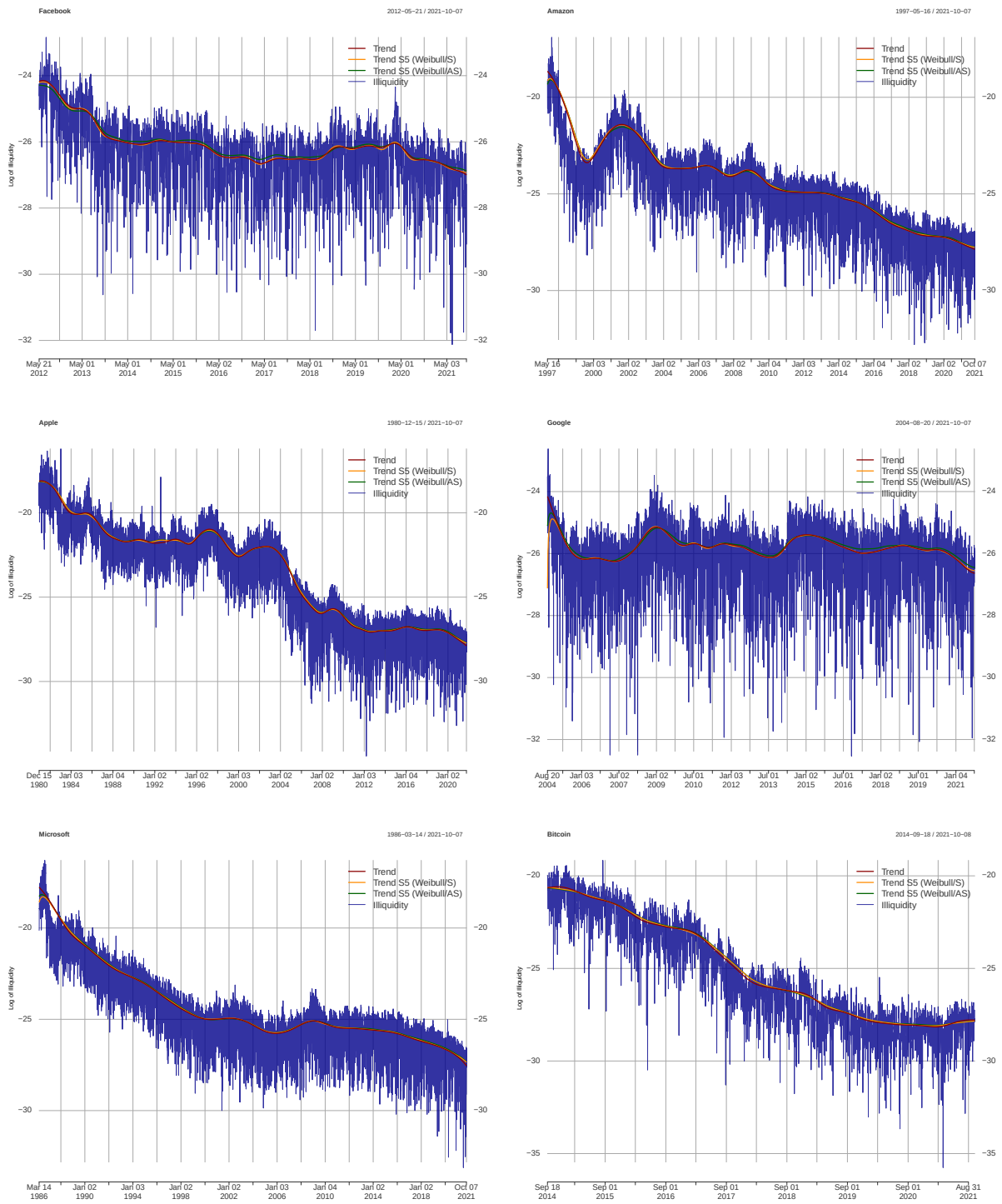


Figure 12: Fab 5 and Bitcoin log illiquidity and updated trend function based on the semiparametric ML estimator of λ_t parameters where the error term ζ_t follows a Weibull distribution. The red curve corresponds to the initial estimate of the trend function and the yellow and green curves correspond to the updated estimates in, respectively, the symmetric and asymmetric specifications of λ_t.

E.5 Estimation: i.i.d. error term with nonparametric density

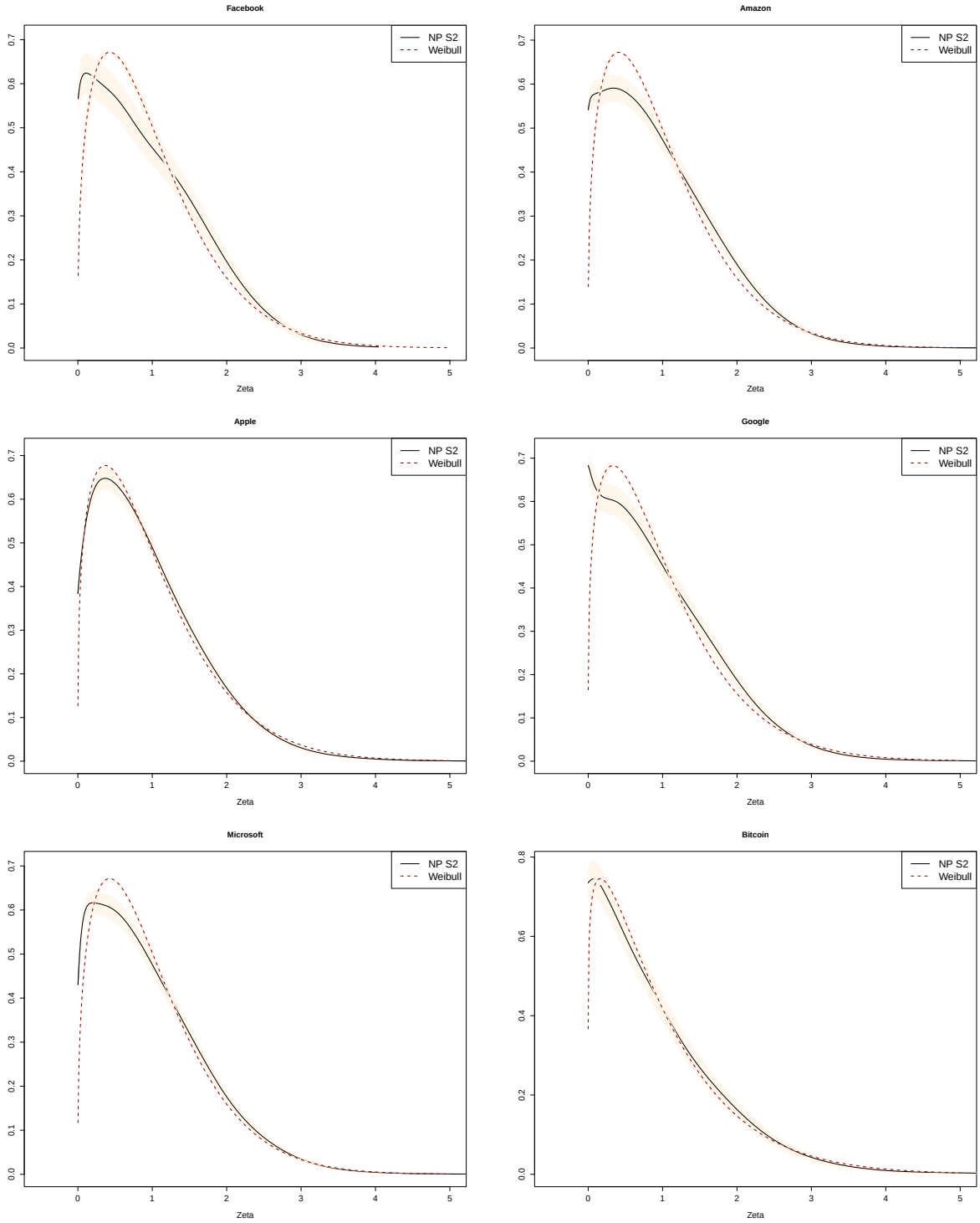


Figure 13: Comparison between the kernel density estimate of ζ_t (solid line) and the Weibull density (dashed line) under the symmetric specification for λ_t.

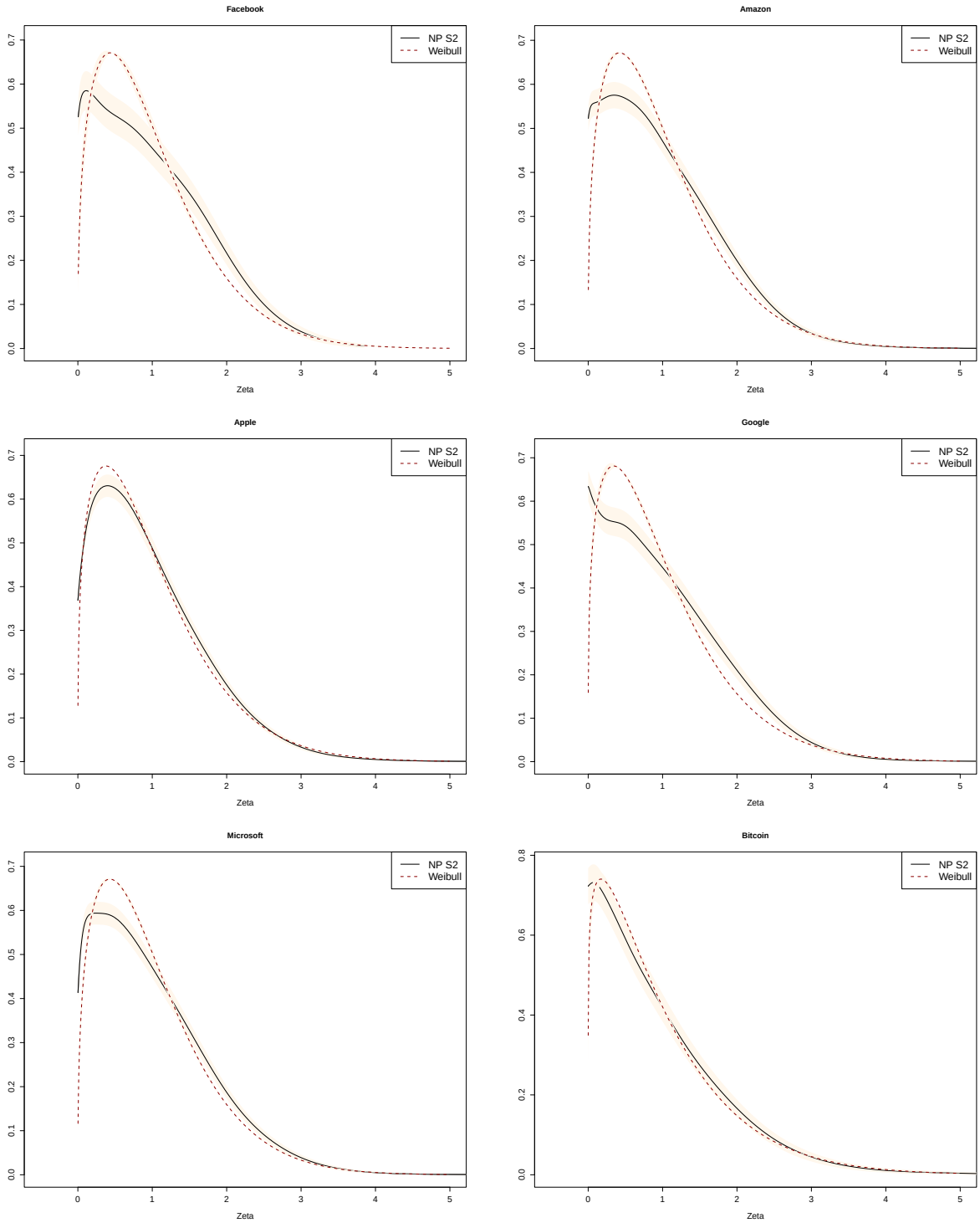


Figure 14: Comparison between the kernel density estimate of ζ_t (solid line) and the Weibull density (dashed line) under the asymmetric specification for λ_t.

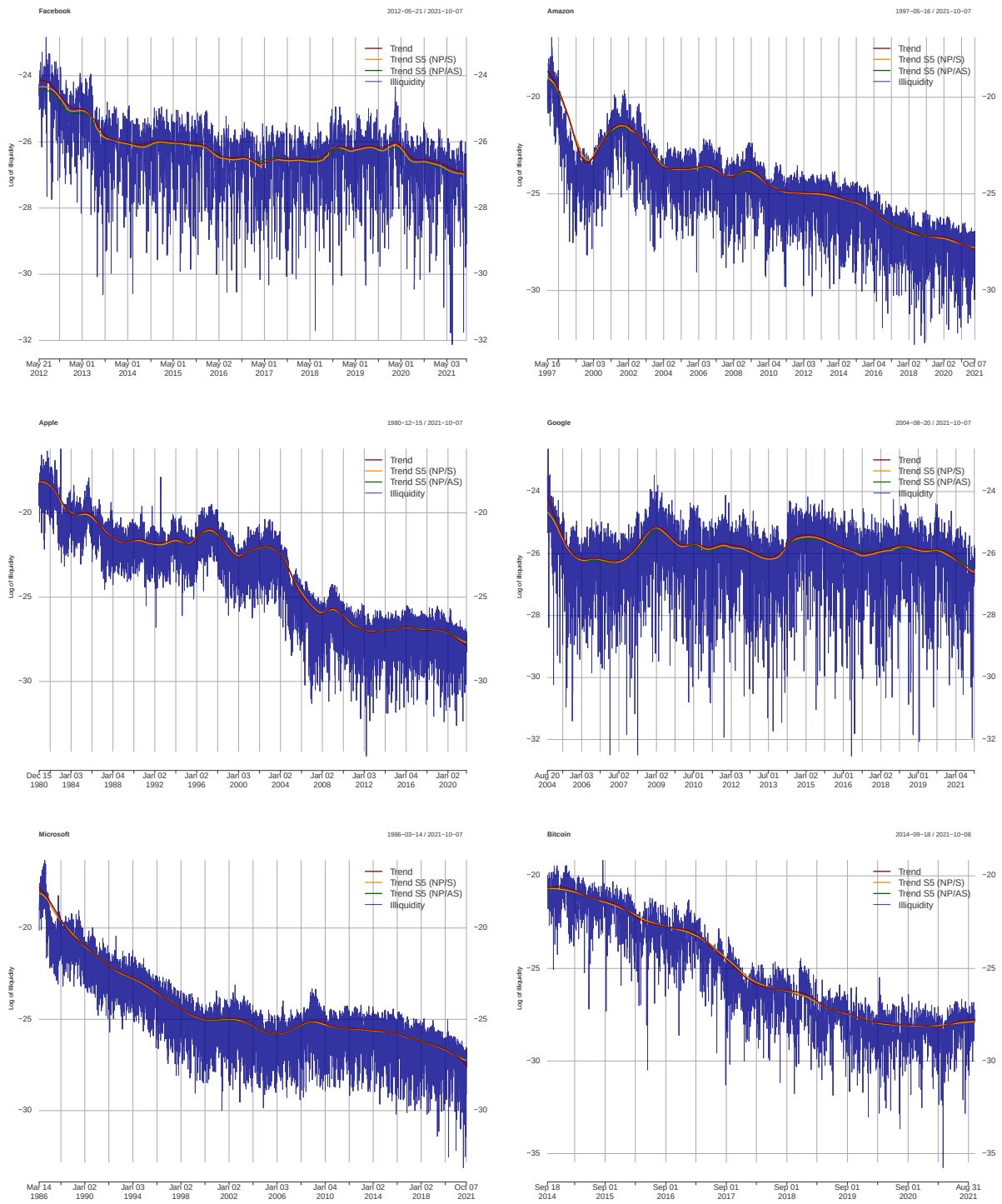


Figure 15: Fab 5 and Bitcoin log illiquidity and updated trend function based on the semiparametric ML estimator of λ_t parameters where the density of the error term ζ_t is estimated nonparametrically. The red curve corresponds to the initial estimate of the trend function and the yellow and green curves correspond to the updated estimates in, respectively, the symmetric and asymmetric specifications of λ_t.

E.6 Testing for permanent shifts: discontinuity in g function

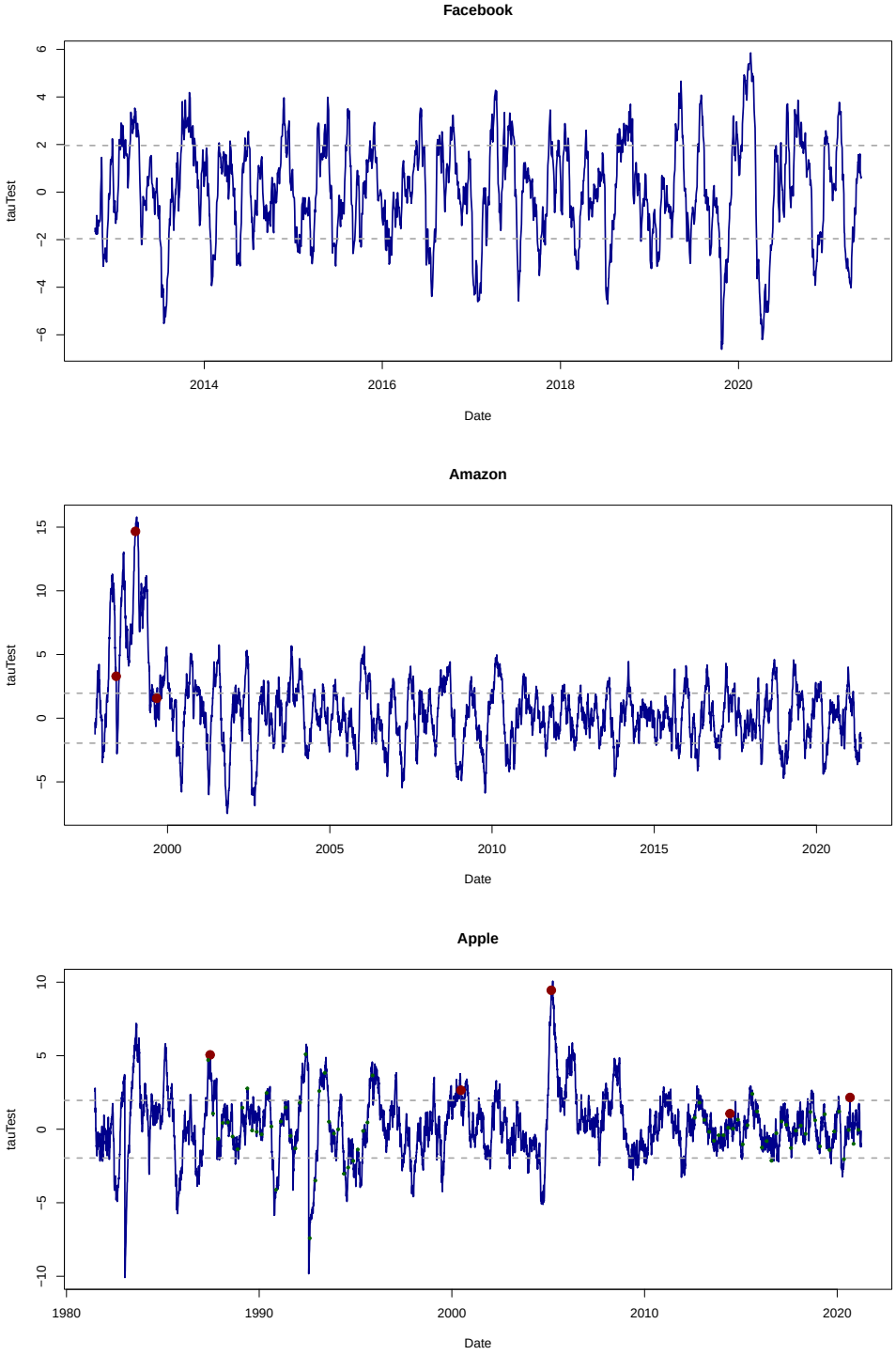


Figure 16: Test statistics for detecting permanent breaks in the illiquidity series.

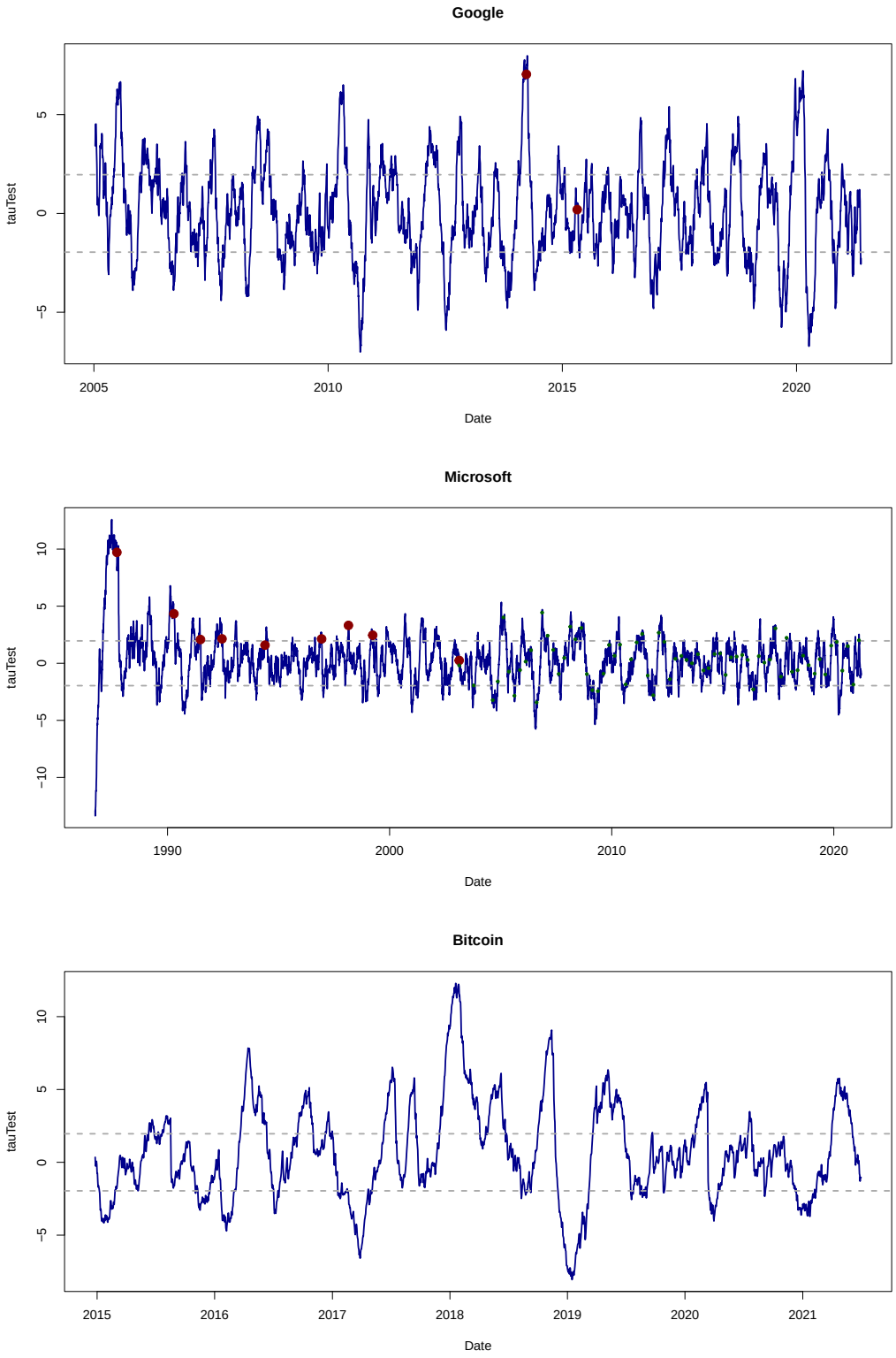


Figure 17: Test statistics for detecting permanent breaks in the illiquidity series.

Table 2: Test statistics for detecting permanent and temporary breaks in the liquidity series of Amazon and Apple stocks.

		Split 1	Split 2	Split 3		
Amazon	Date	1998-06-02	1999-01-05	1999-09-02		
	Splits	02:01	03:01	02:01		
	τ_{LR}	3.29	14.67	1.56		
	τ_{LR}^{AAPL}	2.92	1.33	1.71		
	Δ_{LR}^{3}	-4.18	10.73	-3.38		
		Split 1	Split 2	Split 3	Split 4	Split 5
Apple	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}	5.06	2.65	9.45	1.06	2.16
	τ_{LR}^{MSFT}	-16.16	1.53	7.64	1.34	-1.95
	Δ_{LR}^{3}	-0.04	-0.15	0.11	-2.20	-4.23

Note: We report the test statistic values for the permanent breaks (τ_{LR}), the diff-in-diff test statistics and the change in the long-run illiquidity level normalized by the real tick size $\Delta_{LR}^{3} \times 10^{8}$ where $\Delta_{LR}^{3} = \frac{g^{+}(u)}{\frac{\kappa(u)}{P(u)}} - \frac{g^{-}(u)}{\frac{\kappa(u)}{sP(u)}}$.

Table 3: Test statistics for detecting permanent and temporary breaks in the liquidity series of Google and Microsoft stocks.

		Split 1				
Google	Date	2014-03-27				
	Splits	2002:1000				
	τ_{LR}	7.04				
	τ_{LR}^{SMFT}	-0.30				
	Δ_{LR}^{3}	1.87				
		Split 1	Split 2	Split 3	Split 4	Split 5
Microsoft	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}	9.71	4.33	2.08	2.14	1.59
	τ_{LR}^{AAPL}	7.01	6.55	0.03	-4.02	-4.01
	Δ_{LR}^{3}	0.00	-0.15	-0.08	-0.08	-0.16
		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}	2.13	3.32	2.46	0.25	
	τ_{LR}^{AAPL}	1.76	4.04	-0.83	1.86	
	Δ_{LR}^{3}	-0.00	-0.33	-0.55	-2.12	

Note: We report the test statistic values for the permanent breaks (τ_{LR}), the diff-in-diff test statistics and the change in the long-run illiquidity level normalized by the real tick size $\Delta_{LR}^{3} \times 10^{8}$ where $\Delta_{LR}^{3} = \frac{g^{+}(u)}{\frac{\kappa(u)}{P(u)}} - \frac{g^{-}(u)}{\frac{\kappa(u)}{sP(u)}}$.

E.7 Risk premium

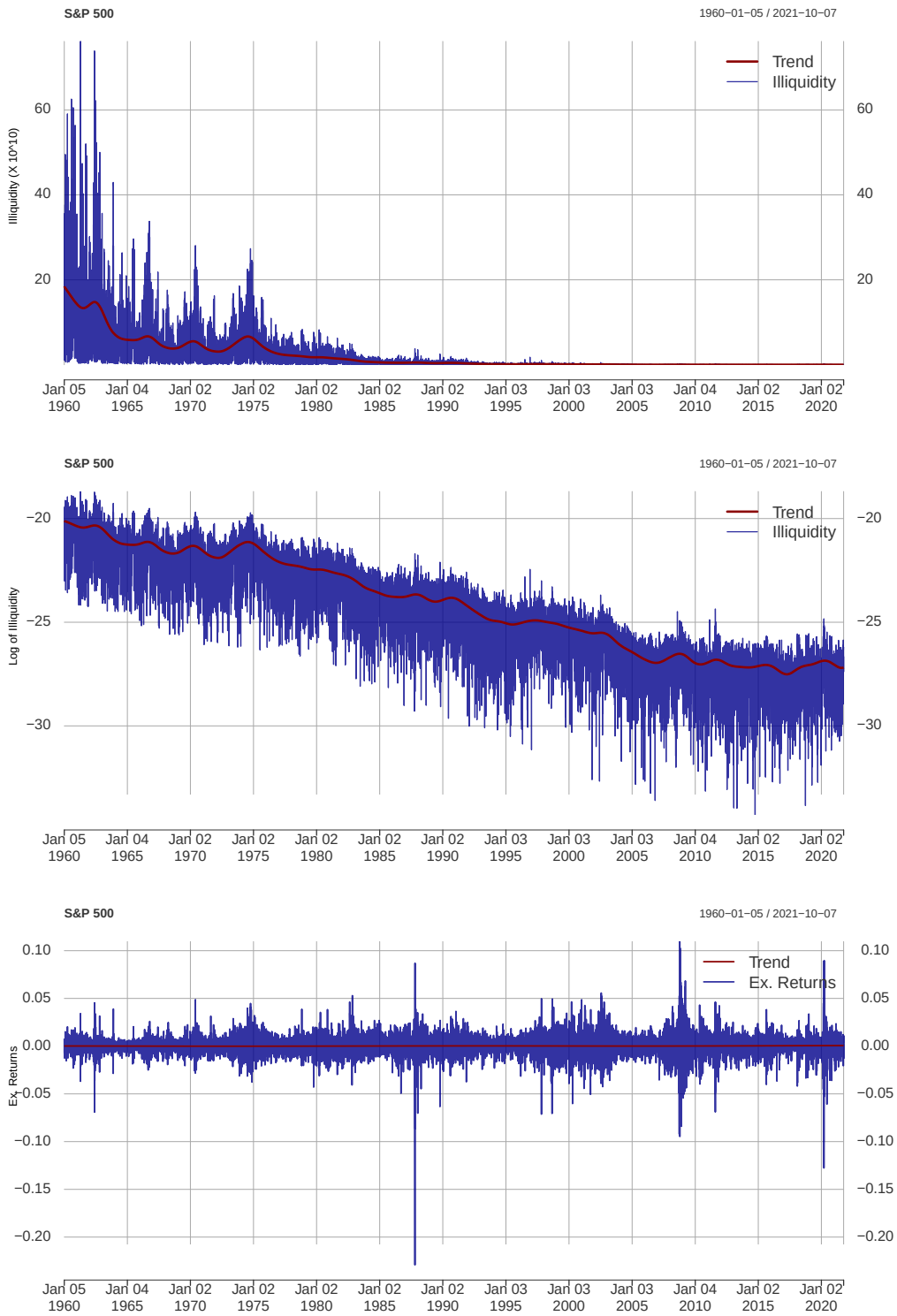


Figure 18: S&P 500 index daily (log) illiquidity series and return data.

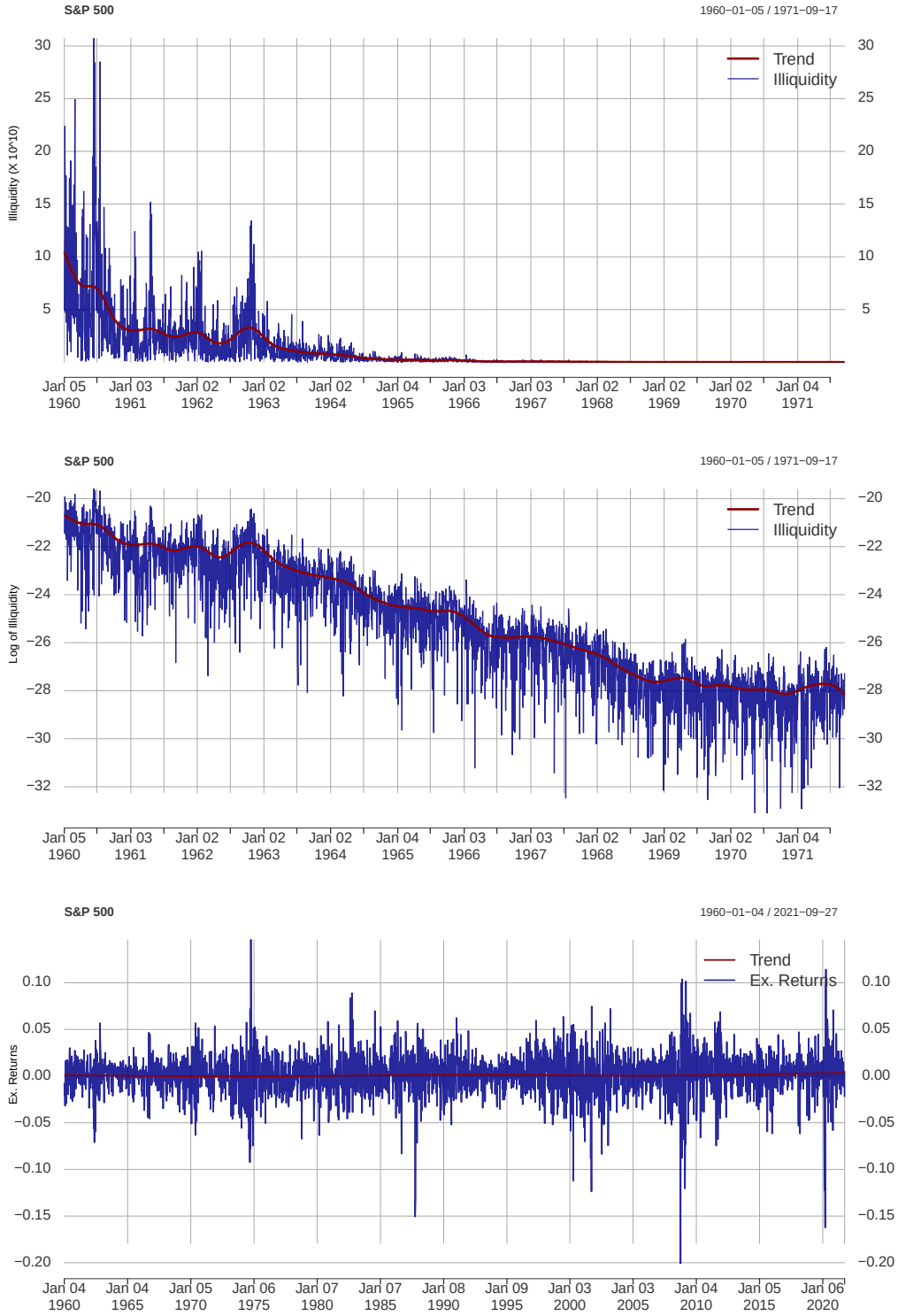


Figure 19: S&P 500 index weekly (log) illiquidity series and return data.

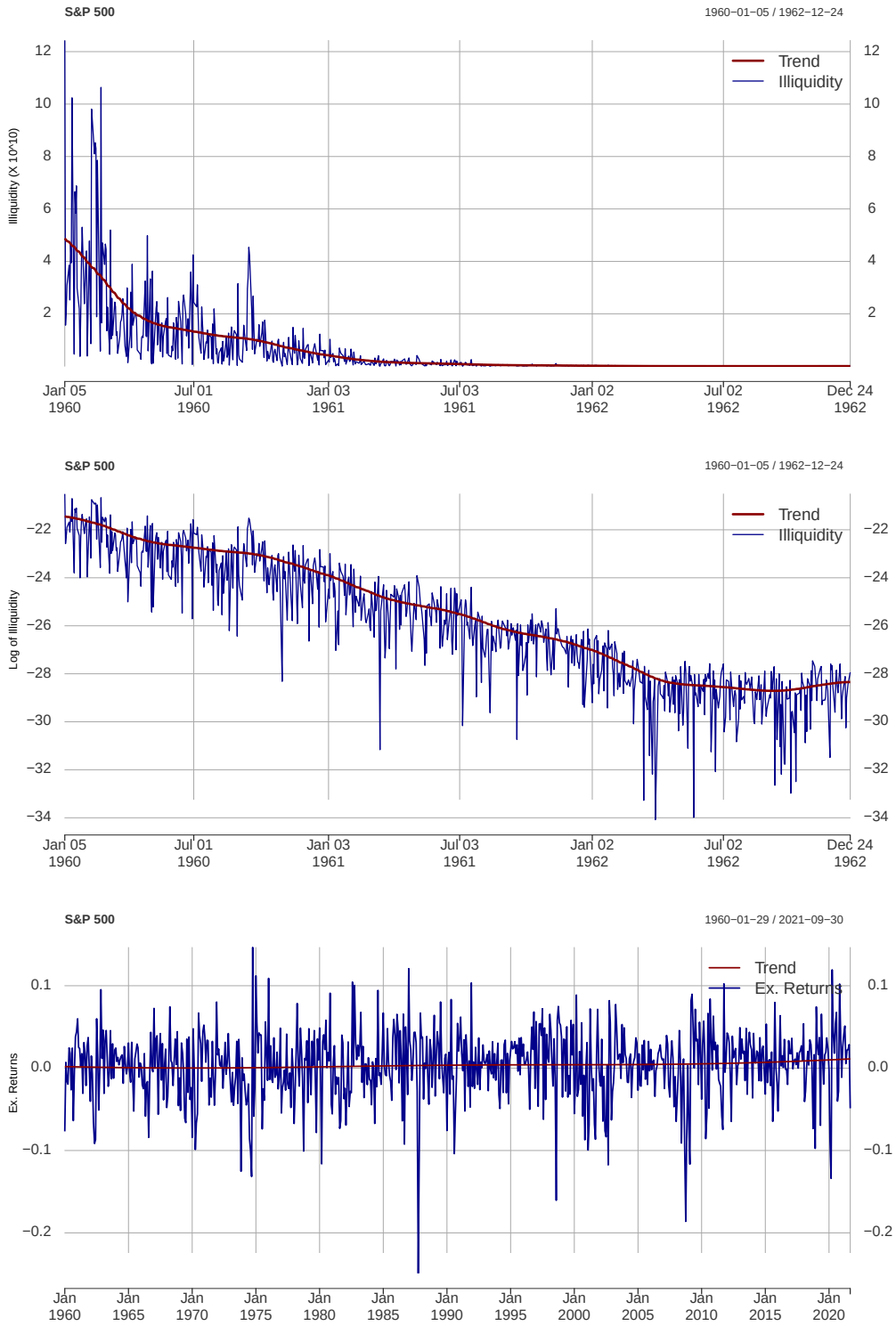


Figure 20: S&P 500 index monthly (log) illiquidity series and return data.

E.8 The occurrence of exact zeros

We consider the daily return data of the S&P 500 stock market index for the period ranging from January 03, 1950 until October 07, 2021. The data contains 125 zero returns in total, which corresponds to 0.69% of the entire sample. To further investigate this issue, we construct a dummy variable which takes a value of one on days where the observed return is zero and plot the resulting series in Figure 21, where we have smoothed the series using a local linear estimator[2]. We denote the smoothed series as $\pi(t/T)$, which is a function of rescaled time representing the unconditional probability of observing a zero at time t. We observe that the zeros series exhibits a strong downward trend over time and the majority of the zeros occurred before 2000. This higher incidence of zero returns in the earlier part of the sample might be linked to low index level (below 100) and restrictions on two decimals for reporting. We further plot in Figure 22 the illiquidity trend function $g(t/T)$ and the corresponding series adjusted for the presence of zero return observations, which we compute as $\frac{g(t/T)}{1-\pi(t/T)}$. It can be observed that there is a small difference between the original estimated trend series and the adjusted one at the beginning of the sample period but the two curves become indistinguishable after 1960.

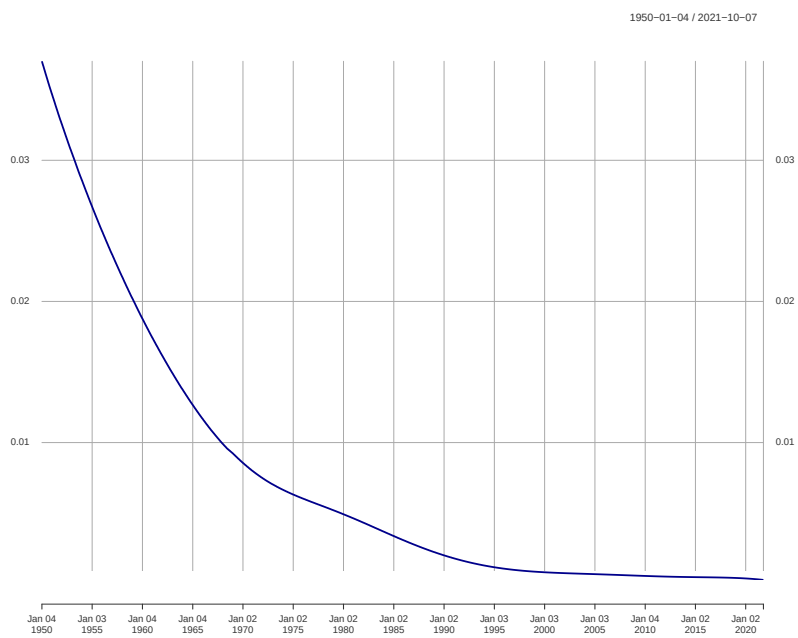


Figure 21: Smoothed series for the occurrence of exact zero returns in the S&P 500 index.

We also compute the ACF of the dummy variable series and plot it in Figure 23 for

[2]We opt for a Gaussian kernel and we choose the bandwidth according to the direct plug-in method as introduced in Ruppert et al. (1995).

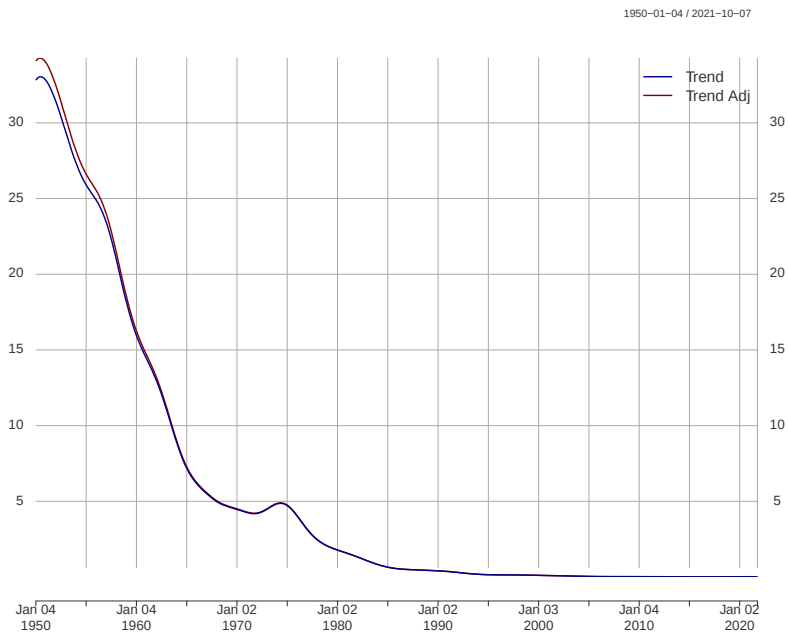


Figure 22: Illiquidity trend function $g(t/T)$ and its corresponding series adjusted for the presence of zero return observations for the S&P 500 index.

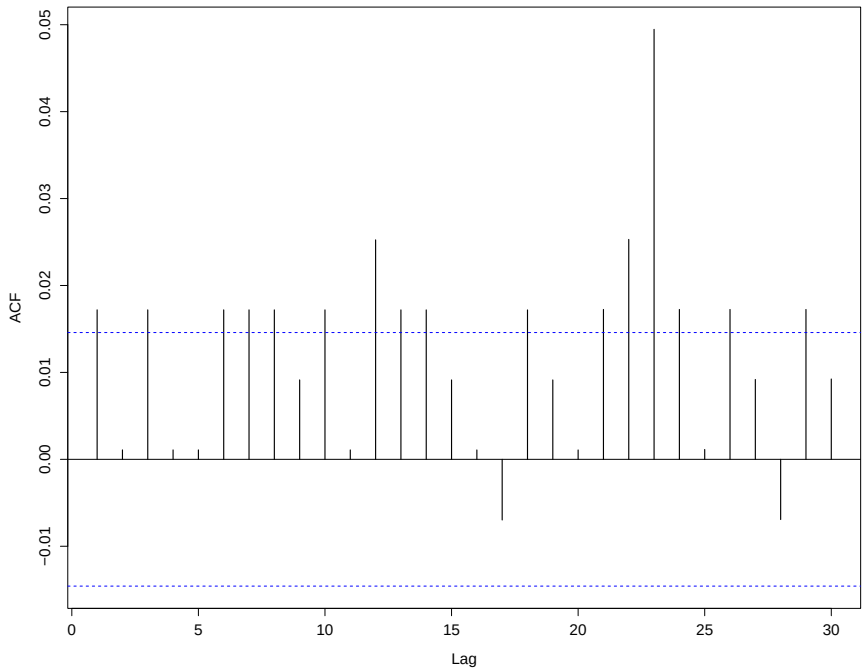


Figure 23: ACF of the series measuring the occurrence of zero return observations in the S&P 500 index.

lags up to 30. The majority of the autocorrelations are positive, with a significant peak at lag 23, and only two of them are negative (lags 17 and 28). In addition, the magnitude of almost all autocorrelations is quite small which might be explained by the relatively infrequent incidence of zero return observations.

E.9 Tail index and fat-tailed distribution

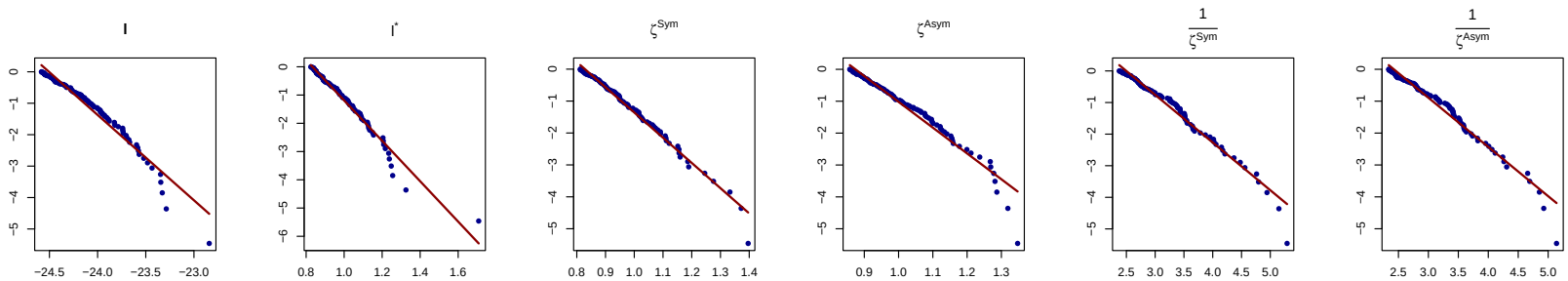
E.9.1 Tail index estimation

We consider the improved tail index estimator $\hat{b}$ proposed by Gabaix and Ibragimov (2011), which is estimated from the regression $\log(\text{Rank} - 1/2) = a - b\log(\text{Size})$ using the 5% of largest observations in the distribution. We report in Table 4 the tail index estimates for the illiquidity ℓ_t, rescaled illiquidity ℓ_t^*, error term ζ_t and its reciprocal $\frac{1}{\zeta_t}$. We observe that the estimated tail index of the error term ζ_t is between three and four for Apple and Bitcoin while it is between four and five for the S&P 500 index. For Facebook, Amazon, Google and Microsoft, the estimated tail index is between six and eight. This suggests that the shocks have a thicker tail than the Weibull distribution. We therefore also consider fat-tailed distributions in our analysis, such as the Lomax, Burr and Inverse Burr distribution.

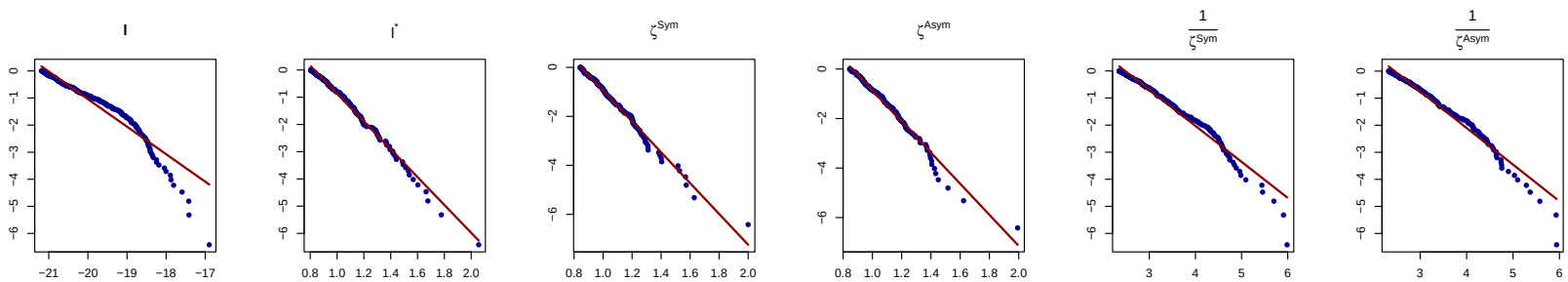
Table 4: Estimated tail index.

	ℓ_t	ℓ_t^*	ζ^{Sym}	ζ^{Asym}	$\frac{1}{\zeta^{\text{Sym}}}$	$\frac{1}{\zeta^{\text{Asym}}}$
S&P500	2.989	4.378	4.902	4.584	1.253	1.244
	(0.14)	(0.21)	(0.23)	(0.22)	(0.06)	(0.06)
Facebook	2.716	7.148	7.906	8.093	1.509	1.543
	(0.35)	(0.93)	(1.03)	(1.05)	(0.20)	(0.20)
Amazon	1.016	5.111	6.293	6.278	1.330	1.355
	(0.08)	(0.41)	(0.51)	(0.51)	(0.11)	(0.11)
Apple	1.331	3.573	3.804	3.836	1.294	1.313
	(0.08)	(0.22)	(0.24)	(0.24)	(0.08)	(0.08)
Google	3.016	6.025	6.450	6.452	1.086	1.086
	(0.29)	(0.58)	(0.62)	(0.62)	(0.10)	(0.10)
Microsoft	0.912	5.141	6.060	6.141	1.738	1.734
	(0.06)	(0.34)	(0.40)	(0.41)	(0.12)	(0.12)
Bitcoin	2.760	4.291	3.695	3.862	0.981	0.983
	(0.35)	(0.54)	(0.46)	(0.48)	(0.12)	(0.12)

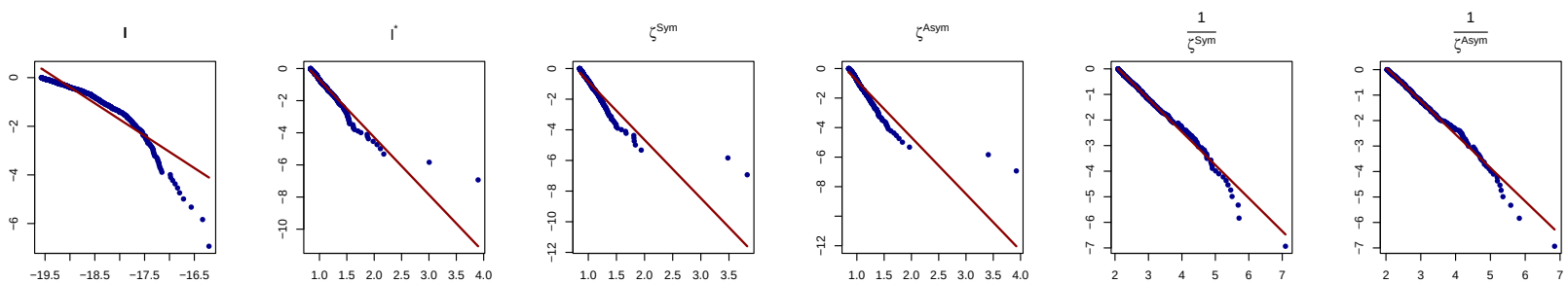
Note: The numbers in the parenthesis are the estimated asymptotic standard errors $(2/n)^{1/2}\hat{b}$ where $\hat{b}$ is the estimated tail index from the regression $\log(\text{Rank} - 1/2) = a - b\log(\text{Size})$. Sym and Asym indicate respectively the symmetric and asymmetric model specifications for λ_t. The regression is based on the 5% percent of largest observations in the distribution.



(a) Facebook

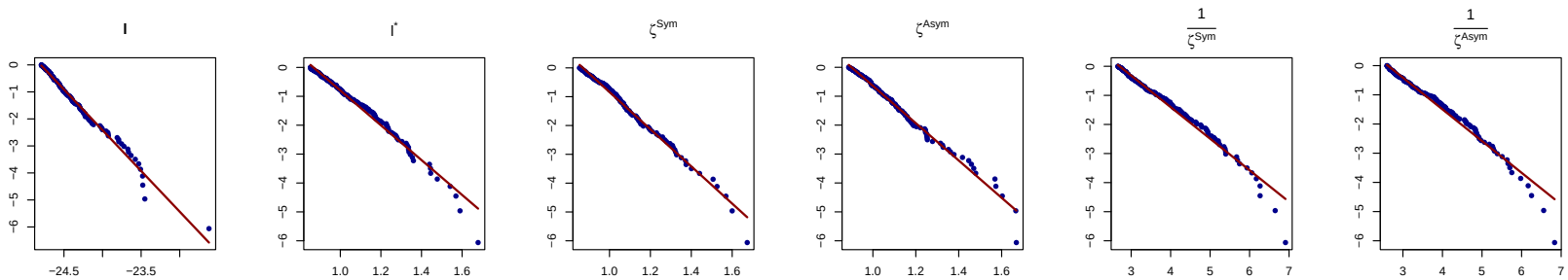


(b) Amazon

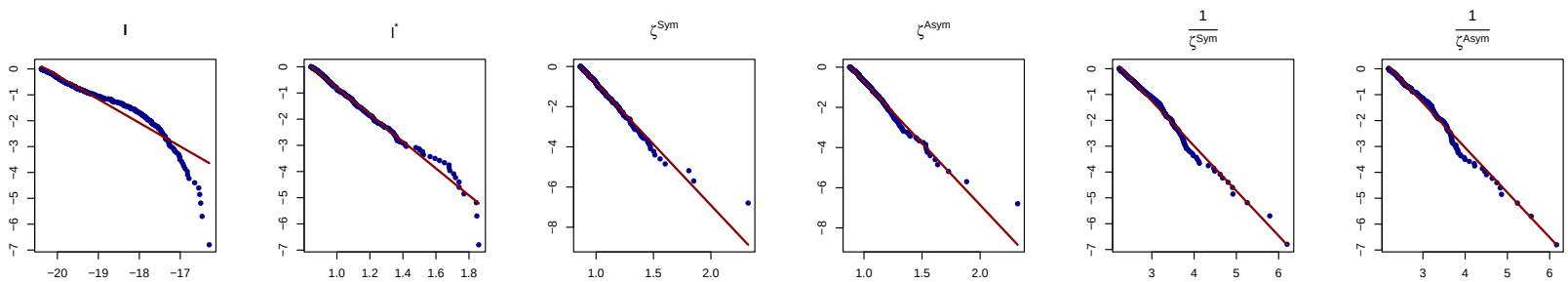


(c) Apple

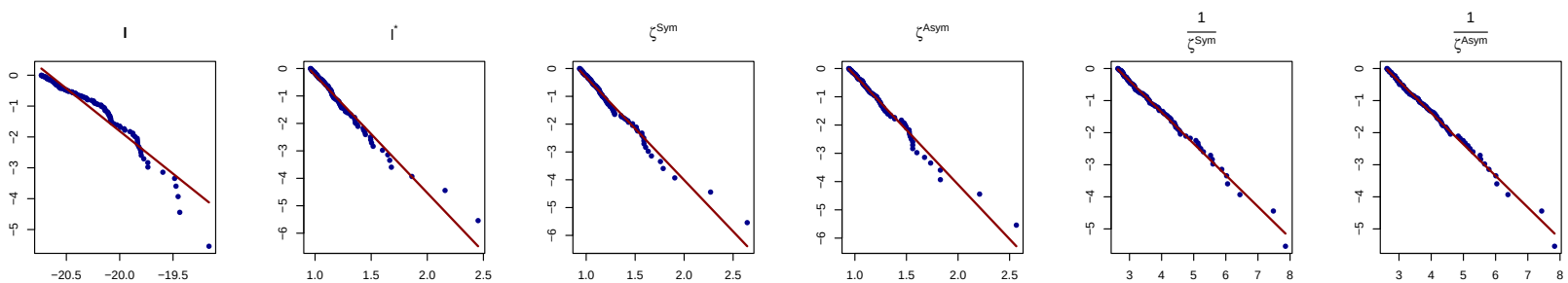
Figure 24: Log(Size) v.s. log((Rank - 1/2)/(n - 1/2)). The slope of the graph corresponds to the estimate of the slope in regression $\log(\text{Rank} - 1/2) = a - b\log(\text{Size})$.



(a) Google



(b) Microsoft



(c) Bitcoin

Figure 25: Log(Size) v.s. log((Rank - 1/2)/(n - 1/2)). The slope of the graph corresponds to the estimate of the slope in regression $\log(\text{Rank} - 1/2) = a - b\log(\text{Size})$.

E.9.2 Maximum likelihood estimation (Weibull, Lomax, Burr and Inverse Burr distributions)

We define the Lomax density function for the random variable $x > 0$, with parameters $\alpha > 0$ and $\lambda > 0$ as

$$f_L(x) = \frac{\alpha}{\lambda}\left[1 + \frac{x}{\lambda}\right]^{-(\alpha+1)}$$

Its uncentered moments of order p are given by

$$\mu_p^L = \frac{\lambda^p \Gamma(\alpha - p)\Gamma(1+p)}{\Gamma(\alpha)}, \text{ for } \alpha > p$$

so that if λ is chosen as $\lambda = \alpha - 1$ then the random variable x has unit mean, i.e. $\mu_1^L = 1$ for $\alpha > 1$. The corresponding variance is equal to

$$\sigma_L^2 = \frac{\alpha}{\alpha - 2}, \quad \alpha > 2$$

We define the Burr density function for the random variable $x > 0$, with parameters $\gamma > 0$, $\lambda > 0$, and $c > 0$ as

$$f_B(x) = \frac{\gamma}{c}\left(\frac{x}{c}\right)^{\gamma-1}\left[1 + \lambda\left(\frac{x}{c}\right)^{\gamma}\right]^{-\left(1+\lambda^{-1}\right)}$$

i.e. $x \sim \text{Burr}(\gamma, \lambda, c)$. Its uncentered moments of order p are given by

$$\mu_p^B = c^p \frac{\Gamma\left(1 + p\gamma^{-1}\right)\Gamma\left(\lambda^{-1} - p\gamma^{-1}\right)}{\lambda^{1+p\gamma^{-1}}\Gamma\left(1 + \lambda^{-1}\right)}, \text{ for } \gamma/\lambda > p$$

so that if c is chosen as

$$c = \frac{\lambda^{1+\gamma^{-1}}\Gamma\left(1 + \lambda^{-1}\right)}{\Gamma\left(1 + \gamma^{-1}\right)\Gamma\left(\lambda^{-1} - \gamma^{-1}\right)}$$

then the random variable x has unit mean, i.e. $\mu_1^B = 1$ for $\gamma > \lambda$. The corresponding variance is equal to

$$\sigma_B^2 = \lambda\Gamma\left(1 + \lambda^{-1}\right)\frac{\Gamma\left(1 + 2\gamma^{-1}\right)\Gamma\left(\lambda^{-1} - 2\gamma^{-1}\right)}{\left[\Gamma\left(1 + \gamma^{-1}\right)\Gamma\left(\lambda^{-1} - \gamma^{-1}\right)\right]^2} - 1$$

- $\lambda \to 0$, the Burr density tends to the Weibull density $W(\gamma, c)$. We note that γ is the shape parameter which we denoted as φ in the main text.
- $\gamma = 1$, the Burr distribution reduces to the Lomax distribution $L(\frac{1}{\lambda}, \frac{c}{\lambda})$.

We define the Inverse Burr density function for the random variable $x > 0$, with parameters $\alpha > 0$, $\theta > 0$, and $\tau > 0$ as

$$f_{IB}(x) = \frac{\alpha\tau(x/\theta)^{\tau\alpha}}{x\left[1 + (x/\theta)^{\tau}\right]^{\alpha+1}}$$

i.e. $x \sim \text{InvBurr}(\alpha, \theta, \tau)$. Its uncentered moments of order p are given by

$$\mu_p^B = \frac{\theta^p \Gamma(1 - p/\tau)\Gamma(\alpha + p/\tau)}{\Gamma(\alpha)}, \text{ for } \tau > p$$

so that if θ is chosen as

$$\theta = \frac{\Gamma(\alpha)}{\Gamma(1 - 1/\tau)\Gamma(\alpha + 1/\tau)}$$

then the random variable x has unit mean, i.e. $\mu_1^{IB} = 1$ for $\tau > 1$. The corresponding variance is equal to

$$\sigma_{IB}^2 = \frac{\Gamma(\alpha)\Gamma(1 - 2/\tau)\Gamma(\alpha + 2/\tau)}{[\Gamma(1 - 1/\tau)\Gamma(\alpha + 1/\tau)]^2} - 1$$

Results summary

We first observe in Table 5 that the Lomax distribution provides an inferior fit compared to the Weibull and Burr distributions. This is due to the fact that when restricting the distribution to have unit mean, the corresponding variance is $a/(a-2) > 1$. However, our data suggests under dispersion with a standard deviation ranging from 0.7 to 0.9.

Secondly, when comparing the results for the Weibull and Burr distributions, we observe that there is a difference in log likelihood of around 170 for Apple and 2 for Bitcoin. For Facebook, Amazon, Google and Microsoft, there is almost no difference in terms of log likelihood. Additionally, the estimated λ^B parameter for the Burr distribution is around 0.09 for Apple, 0.045 for Bitcoin and almost zero for the rest of the stocks which suggests that the estimated Burr distribution reduces to a Weibull distribution (see Table 8). This observation is further confirmed by Figure 26 where we can see that there is a visible difference between the Weibull and Burr distributions for Apple and Bitcoin while for the others the Burr and Weibull densities align with each other.

Lastly, when comparing the results for the Inverse Burr and Burr distributions in the symmetric case, we observe that the Inverse Burr distribution provides a better fit with an increase in log log likelihood ranging from 5 to 50, except in the case of Microsoft whose log likelihood under the Burr distribution is around 30 units larger. From Figure 26, we can observe that the estimated densities for the Inverse Burr distribution depart noticeably from the results obtained with the other distributions. In particular, their behavior around zero requires further investigations.

Table 5: Log likelihood comparison between models using different parametric density (Weibull, Lomax, Burr, Inverse Burr) for the error term ζ_t.

Symmetric specification						
	Facebook	Amazon	Apple	Google	Microsoft	Bitcoin
Weibull	-2171.33	-4938.89	-8843.88	-4019.47	-7440.53	-2397.16
Lomax	-2328.83	-5335.91	-9472.04	-4203.41	-8078.96	-2432.13
Burr	-2171.34	-4938.89	-8674.00	-4019.50	-7440.54	-2394.87
Inv Burr	-2139.68	-4895.35	-8668.71	-3960.89	-7467.99	-2388.48
Asymmetric specification						
Weibull	-2136.45	-4885.79	-8772.52	-3962.31	-7397.12	-2389.26
Lomax	-2309.55	-5309.45	-9430.14	-4165.72	-8051.92	-2426.08
Burr	-2136.49	-4885.93	-8591.43	-3962.33	-7397.13	-2388.32
Inv Burr	-2097.45	-4836.34	-8591.44	-3888.15	-7430.00	-2383.52

Note: The numbers reported are in terms of $\log LL$.

Table 6: Maximum likelihood 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_ζ	β	γ	$\gamma^{(-)}$	φ	σ_ζ
Facebook	0.864	0.047	1.374	0.737	0.904	-0.021	0.098	1.391	0.728
	(38.19)	(5.01)	(73.59)		(86.28)	(-2.36)	(8.94)	(65.46)	
Amazon	0.917	0.079	1.361	0.743	0.928	0.037	0.062	1.372	0.737
	(165.81)	(16.30)	(95.10)		(169.32)	(7.30)	(54.38)	(98.68)	
Apple	0.878	0.095	1.316	0.767	0.891	0.046	0.072	1.321	0.764
	(130.85)	(22.00)	(144.86)		(128.45)	(10.68)	(12.78)	(244.81)	
Google	0.910	0.047	1.279	0.787	0.920	-0.006	0.080	1.296	0.778
	(61.00)	(6.24)	(85.22)		(286.46)	(-2.23)	(28.03)	(460.15)	
Microsoft	0.924	0.068	1.377	0.735	0.926	0.042	0.046	1.381	0.733
	(377.29)	(23.33)	(459.68)		(192.81)	(10.38)	(9.12)	(138.09)	
Bitcoin	0.892	0.063	1.139	0.880	0.893	0.040	0.042	1.144	0.876
	(60.09)	(6.73)	(64.46)		(61.46)	(5.36)	(4.59)	(98.96)	

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 $\sqrt{\frac{\Gamma\left(1+\frac{2}{\varphi}\right)}{\left(\Gamma^2\left(1+\frac{1}{\varphi}\right)\right)} - 1}$. The numbers in parentheses are the t-statistics of the corresponding parameter estimates. We note that the standard errors are under estimated as they do not account for the estimation error associated with the smoothed liquidity process in the first step of the estimation.

Table 7: Maximum likelihood estimates of the parameters for the λ_t process under the assumption that the error term ζ_t follows a Lomax distribution.

	Classic				Asymmetric				
	β	γ	α	σ_ζ	β	γ	$\gamma^{(-)}$	α	σ_ζ
Facebook	0.865	0.047	248.251	1.004	0.905	-0.023	0.098	370.274	1.003
	(8.34)	(0.61)	(1.44)		(33.89)	(-1.74)	(5.17)	(95.48)	
Amazon	0.917	0.078	392.950	1.003	0.927	0.037	0.063	219.623	1.005
	(107.91)	(9.98)	(18.55)		(142.29)	(6.21)	(8.05)	(252.68)	
Apple	0.886	0.090	252.559	1.004	0.899	0.039	0.074	160.438	1.006
	(86.71)	(12.25)	(65.17)		(83.82)	(4.96)	(10.37)	(11.64)	
Google	0.908	0.047	138.007	1.007	0.918	-0.005	0.079	390.310	1.003
	(49.48)	(5.47)	(10.03)		(63.55)	(-0.72)	(6.28)	(27.89)	
Microsoft	0.924	0.069	223.278	1.005	0.927	0.043	0.044	531.198	1.002
	(282.61)	(27.03)	(751482.18)		(139.58)	(7.07)	(7.10)	(12.86)	
Bitcoin	0.898	0.063	145.770	1.007	0.895	0.043	0.039	299.187	1.003
	(64.23)	(7.01)	(4.45)		(58.99)	(4.29)	(3.20)	(3.36)	

Note: The estimated parameters are $\theta = (\beta, \gamma, \alpha)$ for the classic specification and $\theta = (\beta, \gamma, \gamma^{(-)}, \alpha)$ for the asymmetric specification of λ_t. α is the shape parameter of the Lomax distribution which has mean 1 and standard deviation σ_ζ of $\sqrt{\frac{\alpha}{\alpha-2}}, \alpha > 2$. The numbers in parentheses are the t-statistics of the corresponding parameter estimates. We note that the standard errors are under estimated as they do not account for the estimation error associated with the smoothed liquidity process in the first step of the estimation.

Table 8: Maximum likelihood estimates of the parameters for the λ_t process under the assumption that the error term ζ_t follows a Burr distribution.

	Classic					Asymmetric					
	β	γ	γ^B	λ^B	σ_ζ	β	γ	$\gamma^{(-)}$	γ^B	λ^B	σ_ζ
Facebook	0.865	0.047	1.373	0.000	0.737	0.905	-0.023	0.100	1.389	0.000	0.729
	(0.95)	(0.33)	(19.23)	(0.00)		(0.57)	(-0.01)	(0.12)	(9.47)	(0.00)	
Amazon	0.917	0.079	1.361	0.000	0.743	0.928	0.037	0.062	1.376	0.000	0.736
	(0.31)	(0.02)	(0.60)	(0.00)		(5.46)	(0.63)	(0.45)	(1.93)	(0.00)	
Apple	0.891	0.088	1.465	0.092	0.745	0.903	0.040	0.070	1.472	0.087	0.738
	(123.83)	(15.96)	(169.37)	(36.35)		(132.24)	(7.67)	(11.62)	(59.23)	(3.46)	
Google	0.910	0.047	1.281	0.000	0.787	0.920	-0.006	0.079	1.296	0.000	0.778
	(0.05)	(0.01)	(0.11)	(0.00)		(2.33)	(-0.00)	(0.04)	(0.18)	(0.00)	
Microsoft	0.924	0.068	1.377	0.000	0.735	0.925	0.042	0.046	1.381	0.000	0.733
	(7.81)	(0.35)	(19.62)	(0.00)		(0.69)	(0.05)	(0.14)	(0.47)	(0.00)	
Bitcoin	0.895	0.064	1.174	0.045	0.890	0.895	0.042	0.039	1.167	0.029	0.882
	(57.22)	(6.72)	(53.07)	(2.52)		(58.47)	(4.28)	(3.53)	(55.80)	(1.57)	

Note: The estimated parameters are $\theta = (\beta, \gamma, \gamma^B, \lambda^B)$ for the classic specification and $\theta = (\beta, \gamma, \gamma^{(-)}, \gamma^B, \lambda^B)$ for the asymmetric specification of λ_t. (γ^B, λ^B) are the parameters of the Burr distribution which has mean 1 and standard deviation σ_ζ of $\sqrt{\lambda^B \Gamma\left(1+\frac{1}{\lambda^B}\right) \frac{\Gamma\left(1+\frac{2}{\gamma^B}\right)\Gamma\left(\frac{1}{\lambda^B}-\frac{2}{\gamma^B}\right)}{\left[\Gamma\left(1+\frac{1}{\gamma^B}\right)\Gamma\left(\frac{1}{\lambda^B}-\frac{1}{\gamma^B}\right)\right]^2} - 1}$. The numbers in parentheses are the t-statistics of the corresponding parameter estimates. We note that the standard errors are under estimated as they do not account for the estimation error associated with the smoothed liquidity process in the first step of the estimation.

Table 9: Maximum likelihood estimates of the parameters for the λ_t process under the assumption that the error term ζ_t follows an Inverse Burr distribution.

	Classic					Asymmetric					
	β	γ	τ	α	σ_ζ	β	γ	$\gamma^{(-)}$	τ	α	σ_ζ
Facebook	0.866	0.049	6.348	0.146	0.696	0.893	-0.012	0.094	6.658	0.138	0.693
	(29.64)	(5.70)	(30.65)	(25.50)		(68.40)	(-1.57)	(8.74)	(21.29)	(16.89)	
Amazon	0.919	0.076	5.293	0.182	0.723	0.929	0.038	0.058	5.481	0.175	0.715
	(551.23)	(34.23)	(30.91)	(23.97)		(193.85)	(8.43)	(11.14)	(32.21)	(25.75)	
Apple	0.893	0.086	4.177	0.266	0.760	0.901	0.044	0.064	4.290	0.258	0.749
	(124.90)	(16.49)	(51.92)	(35.24)		(138.86)	(8.77)	(12.14)	(41.40)	(29.94)	
Google	0.913	0.046	5.348	0.167	0.749	0.920	-0.003	0.078	5.803	0.150	0.740
	(75.27)	(7.70)	(17.42)	(13.68)		(89.90)	(-0.59)	(8.58)	(24.51)	(19.17)	
Microsoft	0.927	0.066	4.615	0.224	0.741	0.927	0.044	0.040	4.747	0.215	0.736
	(53.00)	(4.82)	(23.73)	(16.44)		(195.94)	(12.38)	(11.63)	(87.81)	(101.75)	
Bitcoin	0.904	0.063	3.601	0.251	0.922	0.903	0.046	0.030	3.678	0.243	0.910
	(67.69)	(7.80)	(27.24)	(21.01)		(62.71)	(5.10)	(3.03)	(42.28)	(25.12)	

Note: The estimated parameters are $\theta = (\beta, \gamma, \tau, \alpha)$ for the classic specification and $\theta = (\beta, \gamma, \gamma^{(-)}, \tau, \alpha)$ for the asymmetric specification of λ_t. (τ, α) are the parameters of the Inverse Burr distribution which has mean 1 and standard deviation σ_ζ of $\sqrt{\frac{\theta^2\Gamma(1-2/\tau)\Gamma(\alpha+2/\tau)}{\Gamma(\alpha)} - 1}$ with $\theta = \frac{\Gamma(\alpha)}{\Gamma(1-1/\tau)\Gamma(\alpha+1/\tau)}$. The numbers in parentheses are the t-statistics of the corresponding parameter estimates. We note that the standard errors are under estimated as they do not account for the estimation error associated with the smoothed liquidity process in the first step of the estimation.

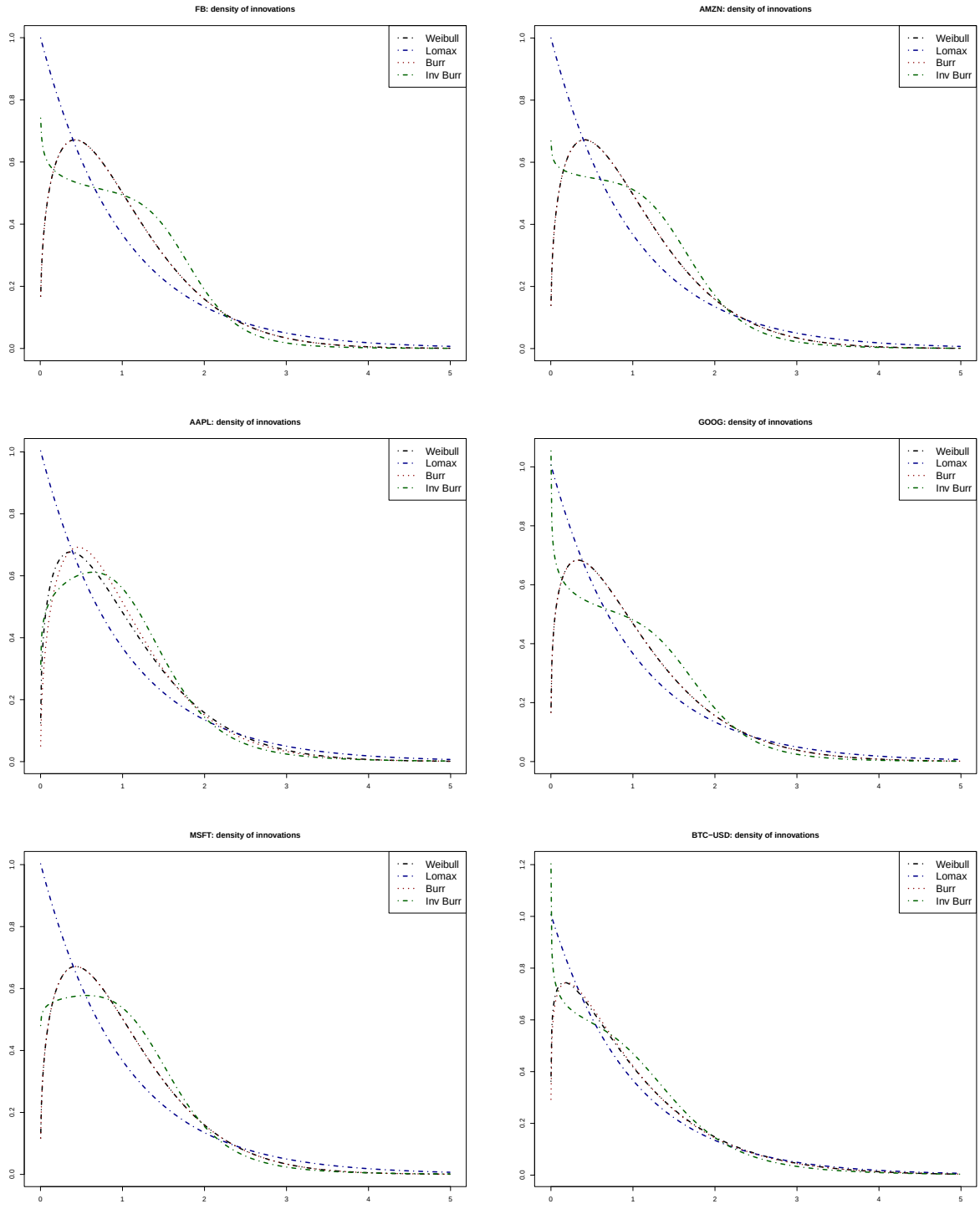


Figure 26: Comparison between the estimated Weibull, Lomax, Burr and Inverse Burr densities under the symmetric specification for λ_t.

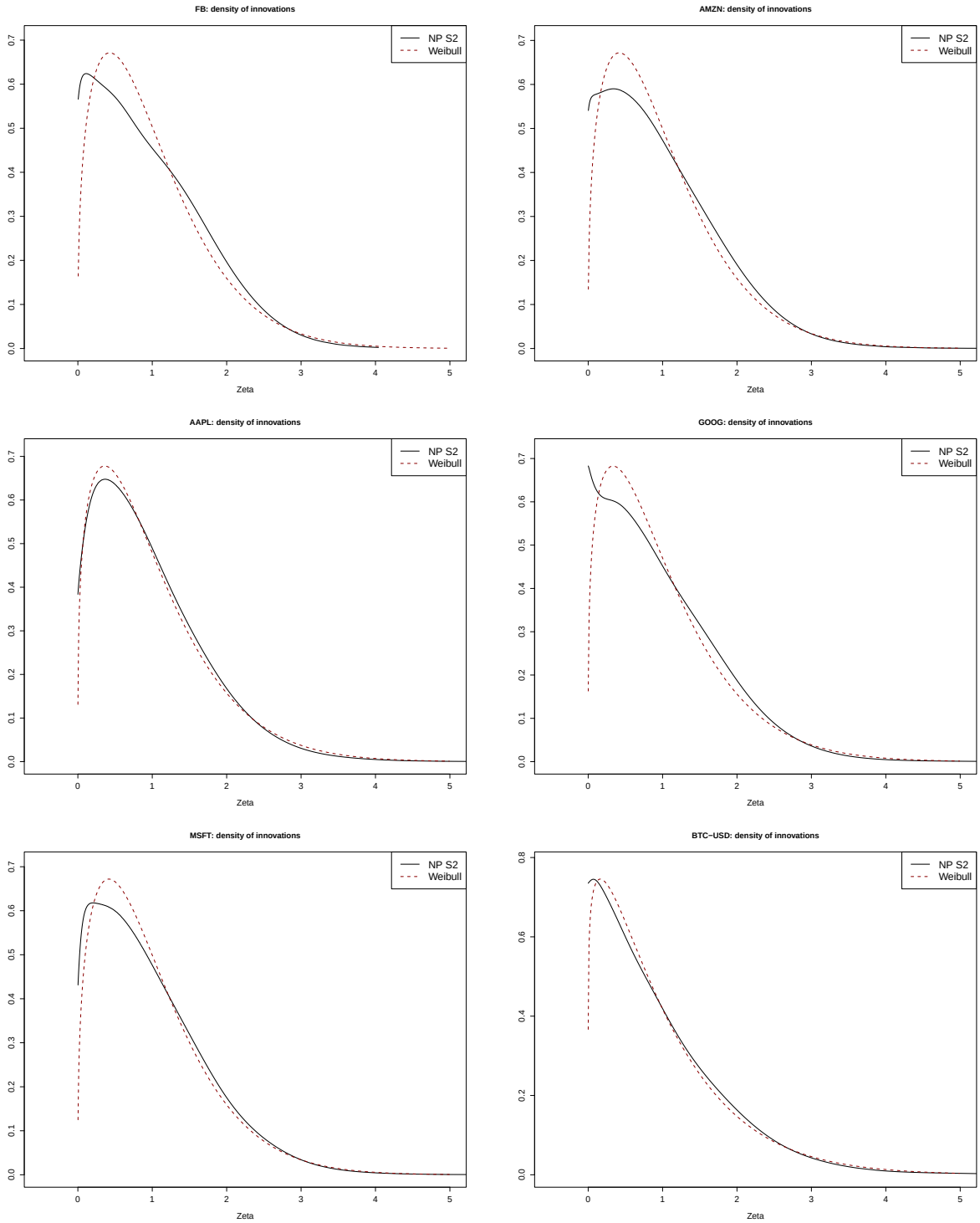


Figure 27: Comparison between the kernel density estimate of ζ_t (solid line) and the Weibull density (dashed line) under the symmetric specification for λ_t.

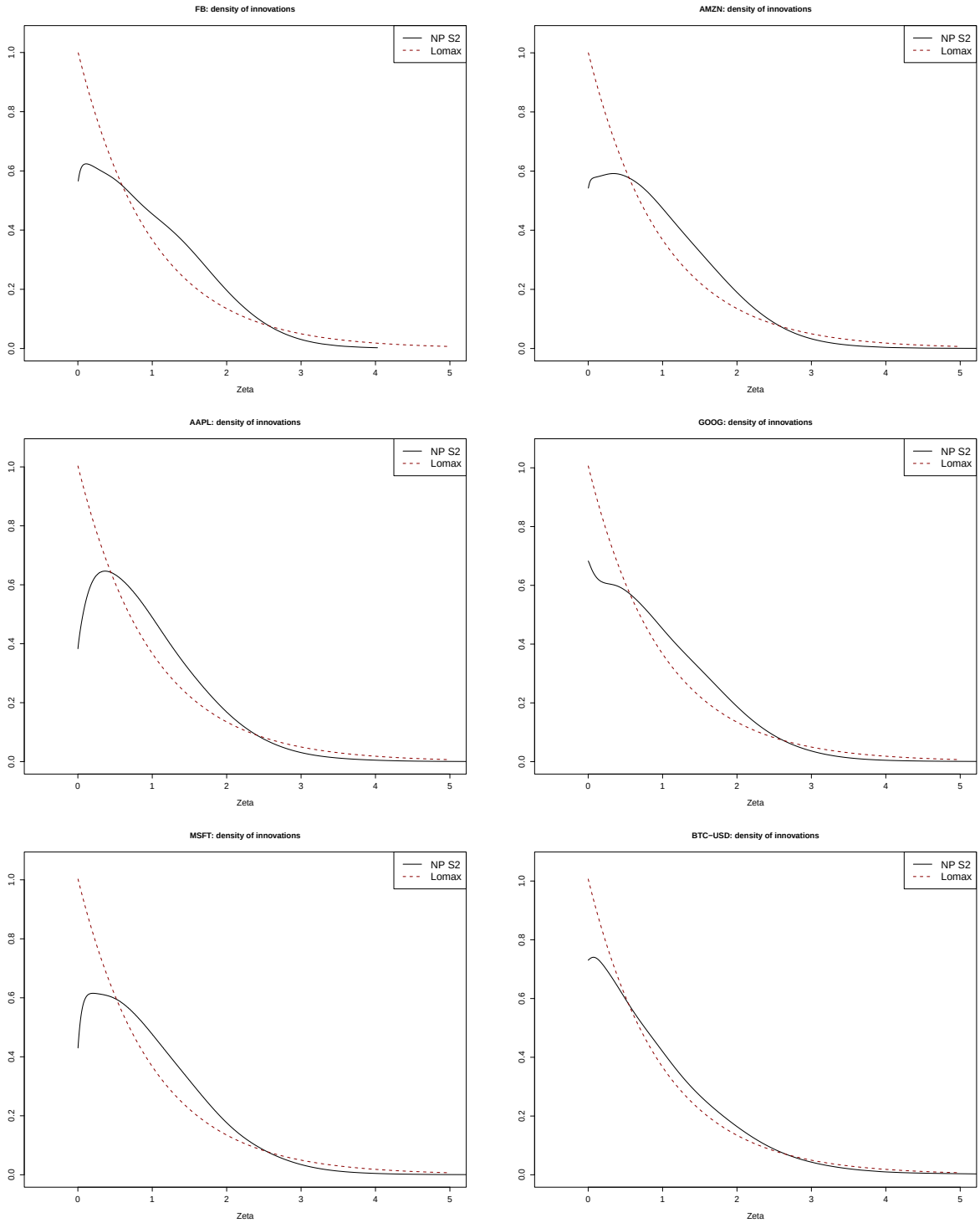


Figure 28: Comparison between the kernel density estimate of ζ_t (solid line) and the Lomax density (dashed line) under the symmetric specification for λ_t.

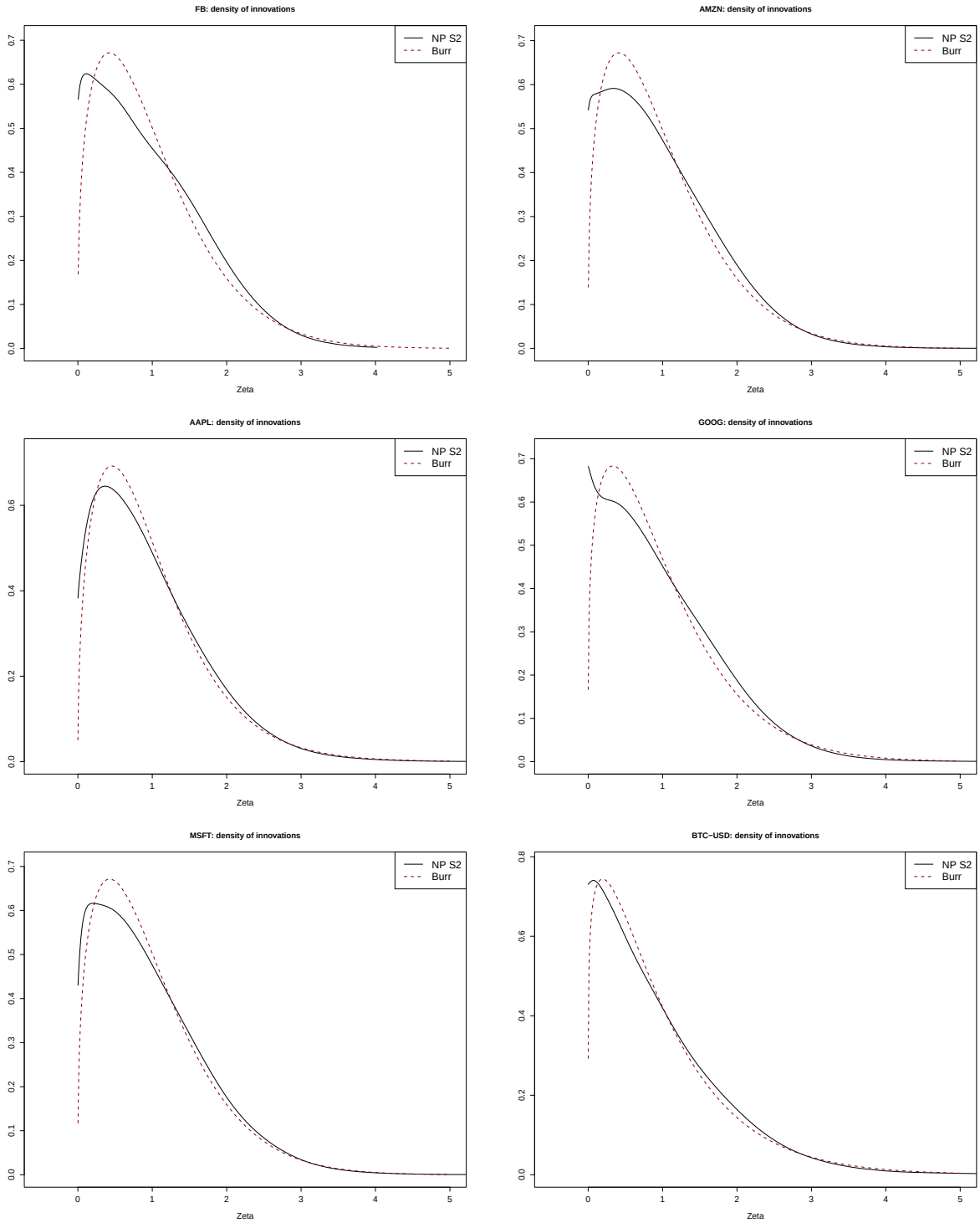


Figure 29: Comparison between the kernel density estimate of ζ_t (solid line) and the Burr density (dashed line) under the symmetric specification for λ_t.

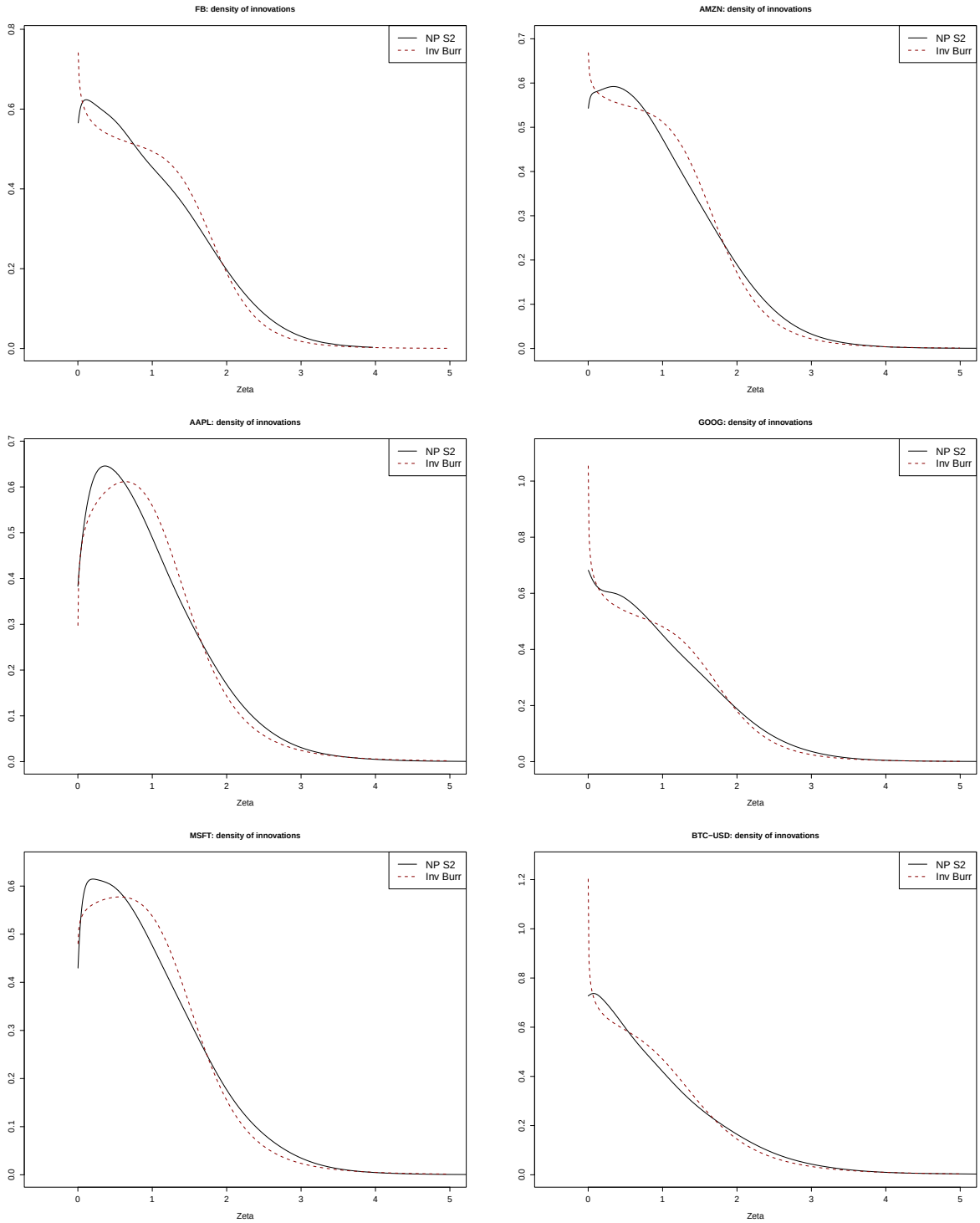


Figure 30: Comparison between the kernel density estimate of ζ_t (solid line) and the Inverse Burr density (dashed line) under the symmetric specification for λ_t.

References

Bickel, P. J., Klaassen, C. A., Ritov, Y. and Wellner, J. A. (1993): *Efficient and adaptive estimation for semiparametric models*. Vol. 4. Springer.

Chen, X., Linton, O. B. and Van Keilegom, I. (2003): "Estimation of semiparametric models when the criterion function is not smooth". *Econometrica* 71.5, pp. 1591–1608.

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

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

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

Francisco-Fernández, M., Vilar-Fernández, J. M. and Vilar-Fernández, J. A. (2003): "On the uniform strong consistency of local polynomial regression under dependence conditions". *Communications in Statistics-Theory and Methods* 32.12, pp. 2415–2440.

Gabaix, X. and Ibragimov, R. (2011): "Rank-1/2: a simple way to improve the OLS estimation of tail exponents". *Journal of Business & Economic Statistics* 29.1, pp. 24–39.

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

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

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

Ruppert, D., Sheather, S. J. and Wand, M. P. (1995): "An effective bandwidth selector for local least squares regression". *Journal of the American Statistical Association* 90.432, pp. 1257–1270.