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

June 28, 2023

Appendices

A Comparison with VLAB implementation

We detail the comparison of our DArLiQ model with the VLAB implementation. 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

*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.

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.

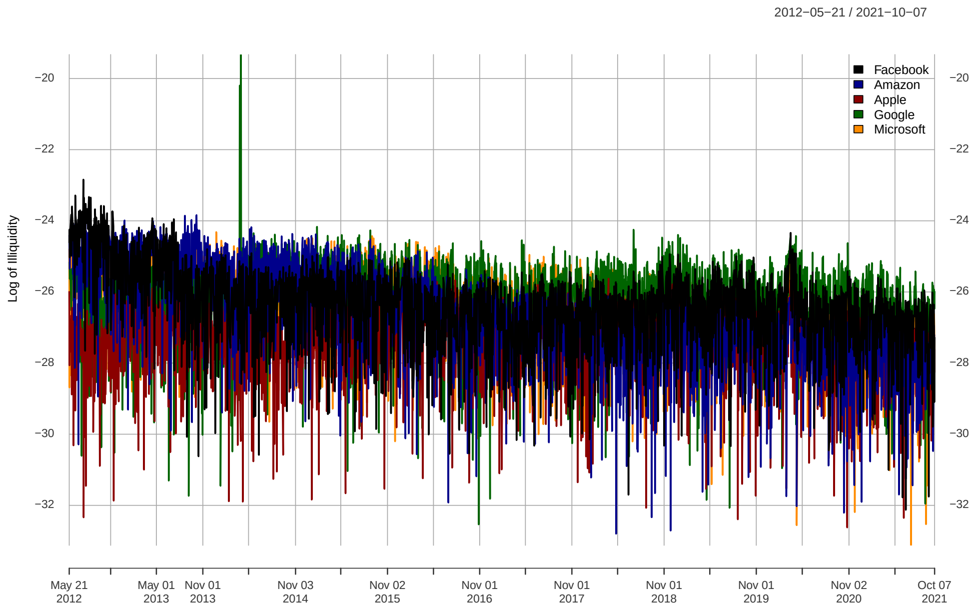
B Amihud illiquidity

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 –

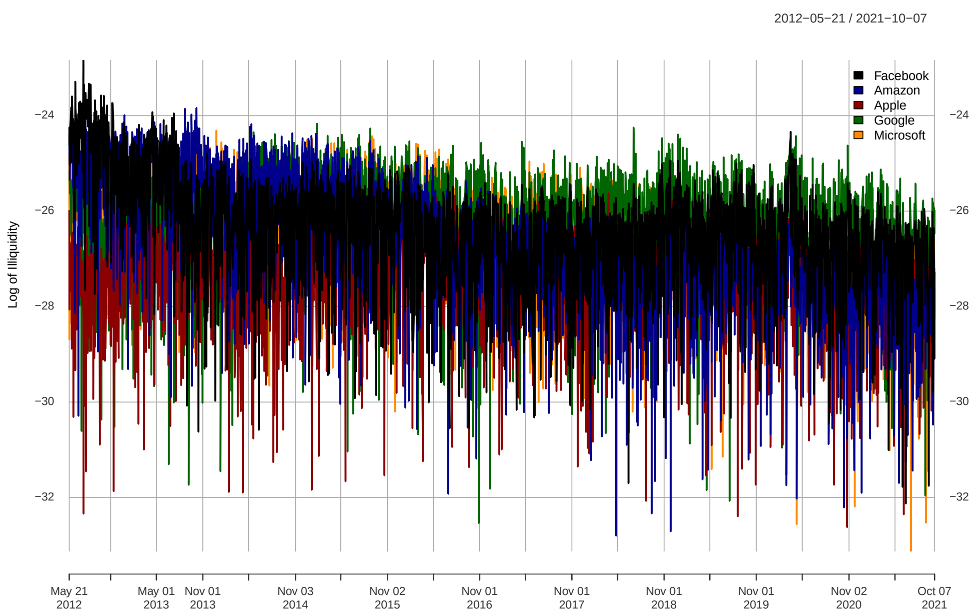
over the period from May 2012 to October 2021. Note that there is a spike in the illiquidity series for Google around end-March 2014 which is caused by a stock split on March 27, 2014.[1] As this event caused irregularity in the trading activities for a few days, we replace the volume data on those dates 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.

To emphasize how prevalent trends in illquidity are 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. 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, 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.

[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.



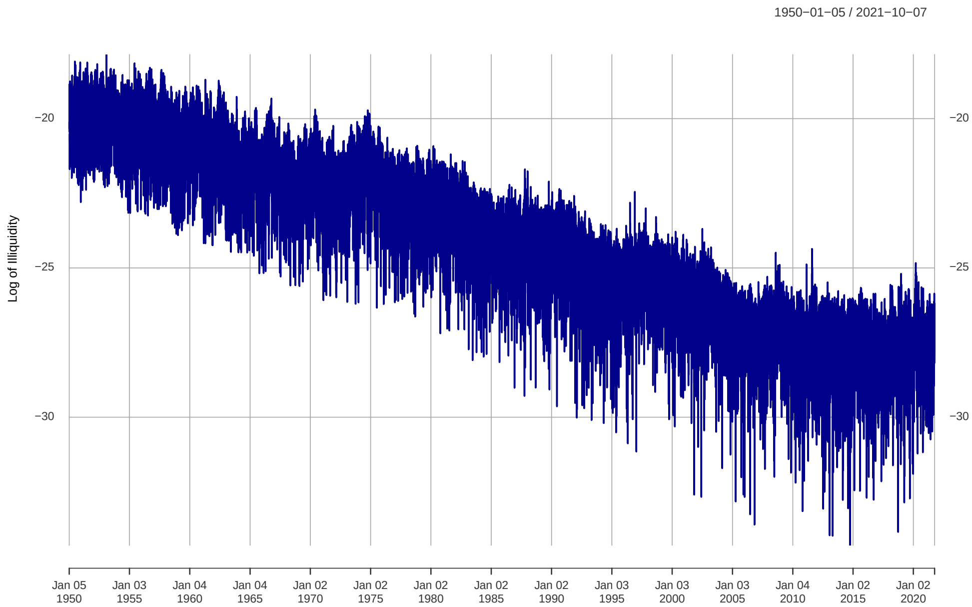
(b) Data of Google stock processed.

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

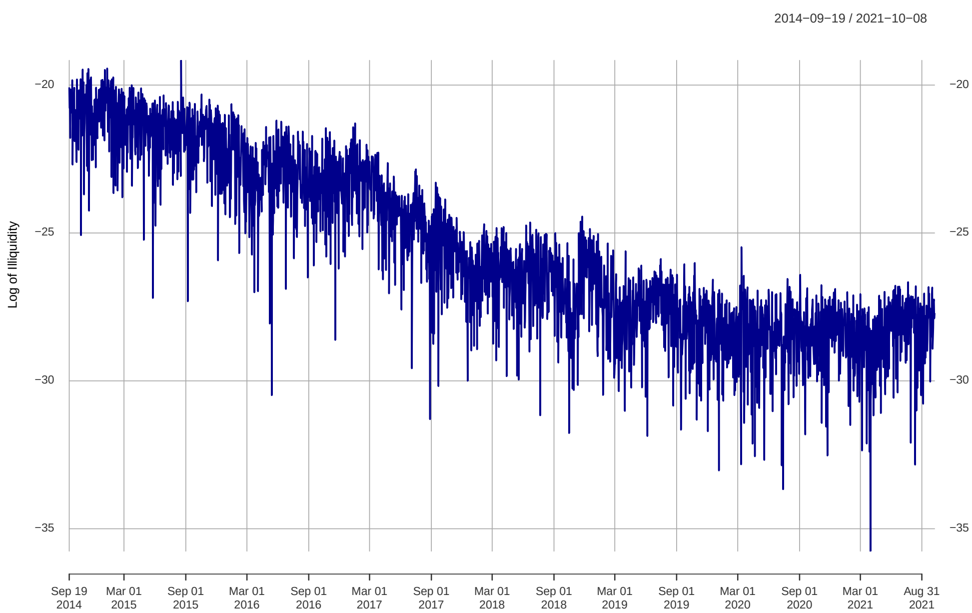
Table 1: Estimated parameters of an AR(5) with trend.

	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 – $\log \ell_t$.

C 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:

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

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

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

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

$$\frac{1}{T}\sum_{t=1}^{T} K_h(t/T-u)g(t/T)v_t = g(u)\frac{1}{T}\sum_{t=1}^{T} K_h(t/T-u)v_t + \frac{1}{T}\sum_{t=1}^{T} K_h(t/T-u)\big(g(t/T)-g(u)\big)v_t$$
$$= V_{T1}(u) + V_{T2}(u).$$

Furthermore,

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

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

$$V_{T2}(u) = hg'(u)\frac{1}{T}\sum_{t=1}^{T} L_{1h}(t/T-u)v_t + h^2 g''(u)\frac{1}{2T}\sum_{t=1}^{T} L_{2h}(t/T-u)v_t$$
$$+ h^2\frac{1}{2T}\sum_{t=1}^{T} L_{2h}(t/T-u)\big(g''(q^*(t/T,u)) - g''(u)\big)v_t,$$

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,

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

This follows because $\mathcal{A} = \left\{\min_{1\le t\le T} \widehat{g}(t/T) < c/2\right\} \subset \mathcal{B} = \left\{\max_{1\le t\le T} \left|\widehat{g}(t/T) - g(t/T)\right| > c/2\right\}$, 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 $\mathbb{R}_+$,

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

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 $\widehat{\zeta}_t - \zeta_t$.

D 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

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

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

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

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

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

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

$$\begin{aligned}
\lambda_t(\widehat{\theta},g_0)-\lambda_t &= \frac{\partial\lambda_t(\theta_0,g_0)}{\partial\theta^{\intercal}}\left(\widehat{\theta}-\theta_0\right)+\left(\widehat{\theta}-\theta_0\right)^{\intercal}\frac{\partial^2\lambda_t(\theta_0,g_0)}{\partial\theta\partial\theta^{\intercal}}\left(\widehat{\theta}-\theta_0\right) \qquad (2)\\
&+\left(\widehat{\theta}-\theta_0\right)^{\intercal}\left(\frac{\partial^2\lambda_t(\overline{\theta},g_0)}{\partial\theta\partial\theta^{\intercal}}-\frac{\partial^2\lambda_t(\overline{\theta},g_0)}{\partial\theta\partial\theta^{\intercal}}\right)\left(\widehat{\theta}-\theta_0\right),
\end{aligned}$$

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

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

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

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

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

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

We have

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

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

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

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

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

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

It follows that

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

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

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

At the true parameter values

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

We have 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

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

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

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

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

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

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

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

For example, consider the following Beta density $\tau(z) = B(a+1)^{-1}z^a(1-z)^a, \quad 0 \leq z \leq 1$, 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 $a \geq 3$. 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 $\theta \in \Theta$.

E Semiparametric efficiency

E.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,

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

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

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

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

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

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

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

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

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

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

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

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

$$= I_2(f)\sum_{t=1}^{T}\left[E\left(\frac{1}{\lambda_t^2}\frac{\partial \lambda_t}{\partial \theta}\right) - \frac{E\left[\frac{\partial \lambda_t}{\partial \theta}\frac{1}{\lambda_t^2}\right]}{E\left(\frac{1}{\lambda_t^2}\right)}E\left(\frac{1}{\lambda_t^2}\right)\right] h(t/T)$$

$$= 0.$$

E.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

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

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

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

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

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

$$= 0$$

for any h.

E.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.

F Nonparametric bootstrap procedure for risk premium regressions

In this section, we summarize the nonparametric bootstrap procedure to obtain standard errors for the risk premium regression under our illiquidity modelling framework. Suppose that

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

$$\overline{R}_t = R_t - m(t/T) = \alpha + \gamma\lambda_t + \delta\zeta_t + e_t.$$

1. Let $\widehat{g}(.), \widehat{\theta}$ be the estimates of the illiquidity model, obtained using the GMM approach or the semiparametric ML estimation procedure. Let $\widehat{\lambda}_t = \lambda_t(\widehat{\theta}), \widehat{\zeta}_t = \ell_t/\widehat{g}(t/T)\widehat{\lambda}_t$.

2. Use the estimated coefficients $\widehat{\alpha}, \widehat{\gamma}, \widehat{\delta}$ to obtain the OLS residuals from the regression of $R_t - \widehat{m}(t/T)$ on $1, \widehat{\lambda}_t, \widehat{\zeta}_t$ and denote $\widehat{e}_t$ as the OLS residuals from the return regression.

3. Let $\widetilde{u}_t = (\widehat{\zeta}_t/\sum_{t=1}^{T}\widehat{\zeta}_t/T, \widehat{e}_t)$ and resample jointly with replacement $\widetilde{u}_t^* = (\zeta_t^*, e_t^*)$, $t = 1, \ldots, T$. Then use ζ_t^* to generate new series of λ_t^* and $\ell_t^* = \widehat{g}(t/T)\lambda_t^*\zeta_t^*$ and then reestimate $\widehat{g}^*(.)$ and $\widehat{\theta}^*$ and hence $\widehat{\lambda}_t^*$, $\widehat{\zeta}_t^*$.

4. Then with e_t^*, we generate new series of $\overline{R}_t^*$

$$\overline{R}_t^* = \widehat{\alpha} + \widehat{\gamma}\widehat{\lambda}_t^* + \widehat{\delta}\widehat{\zeta}_t^* + e_t^*.$$

5. Re-estimate the return regression by OLS to obtain $\widehat{\alpha}^*, \widehat{\beta}^*, \widehat{\gamma}^*$. The standard errors for $\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}$ can be obtained from the distribution of $\widehat{\alpha}^*, \widehat{\beta}^*, \widehat{\gamma}^*$ across resamples.

G Other tables and figures

G.1 Estimation of long-run trend function

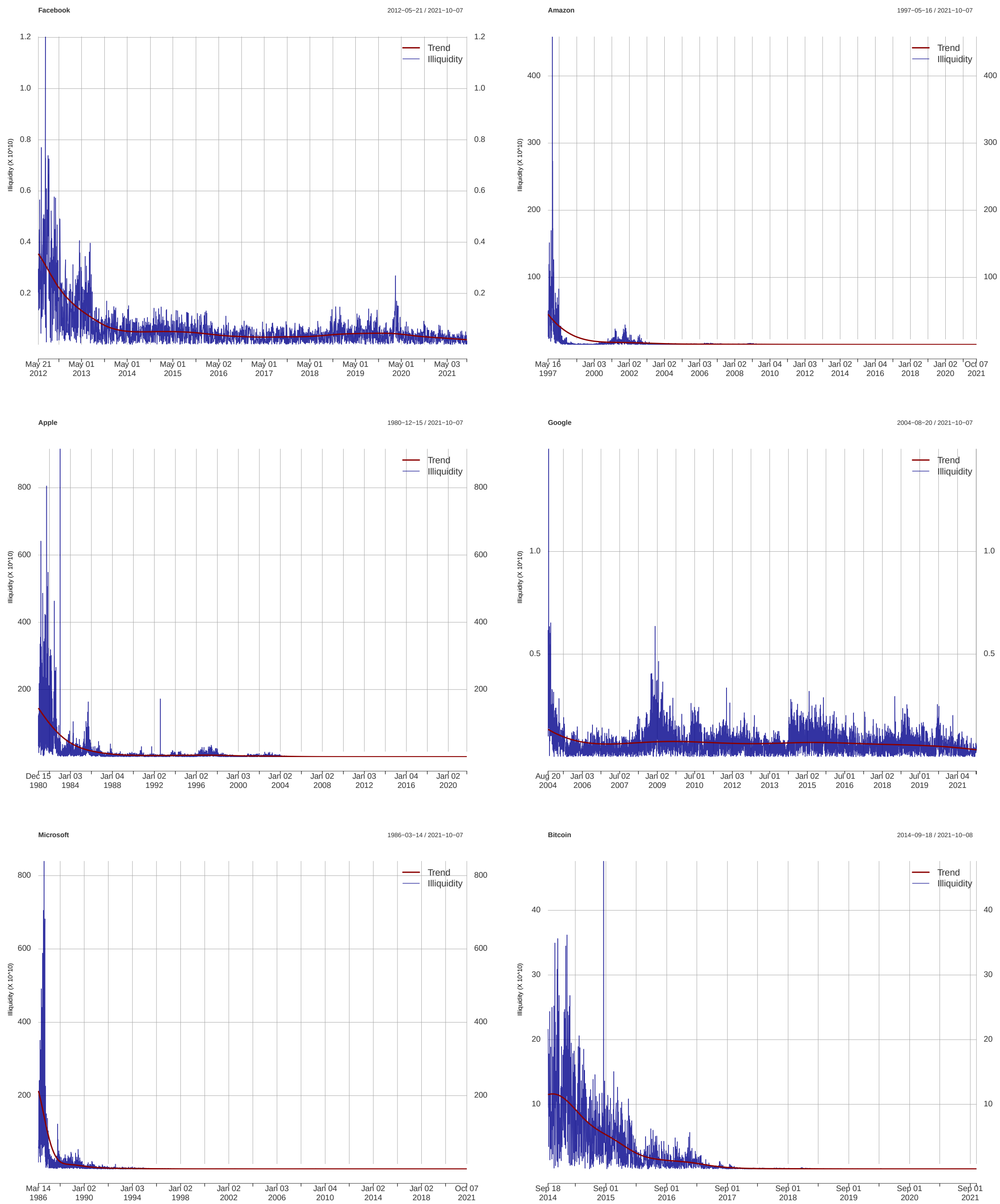


Figure 3: Fab 5 and Bitcoin illiquidity series and trend functions ($\times 10^{10}$).

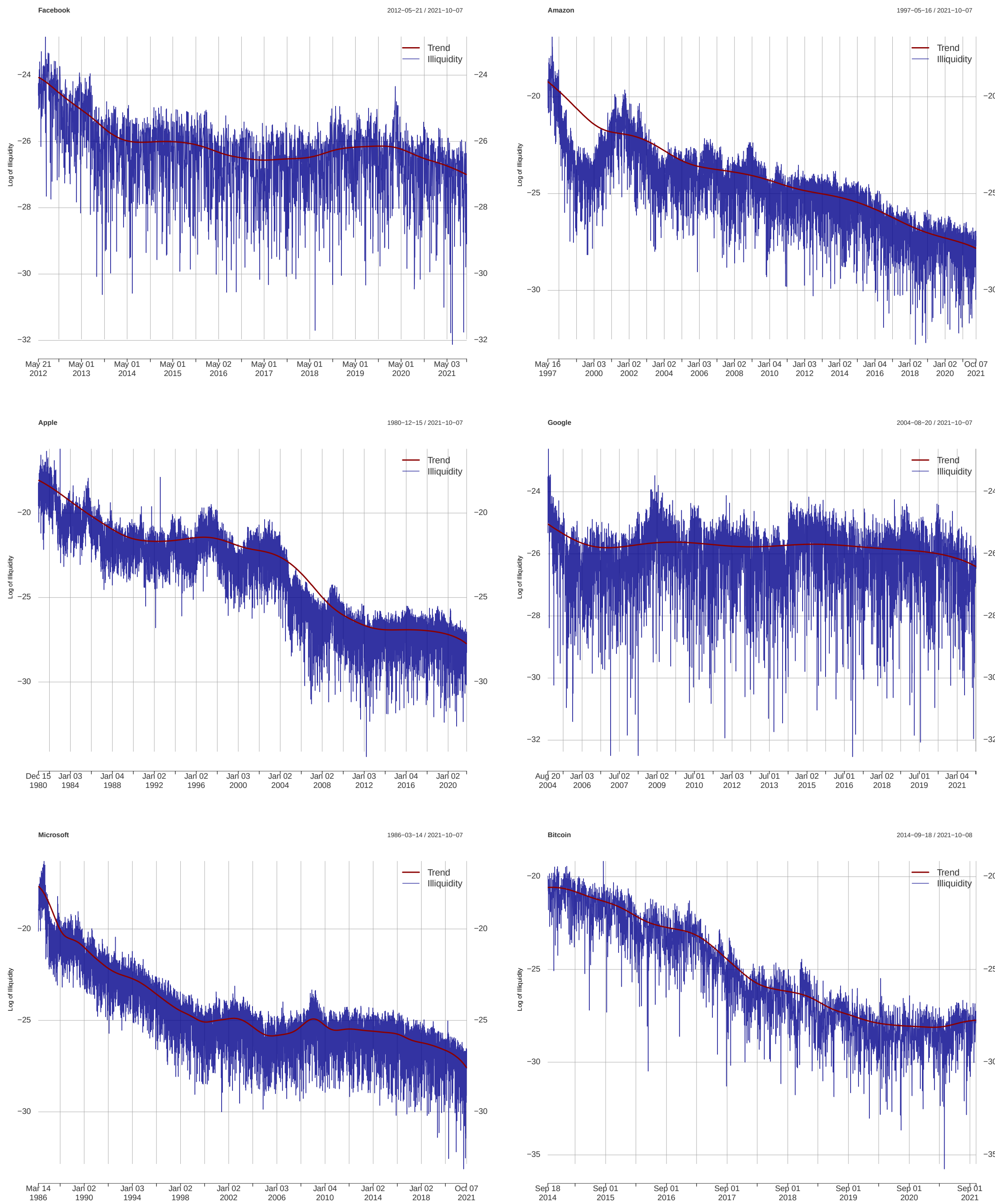


Figure 4: Fab 5 and Bitcoin log illiquidity series and trend functions.

G.2 Estimation based on conditional moment restrictions

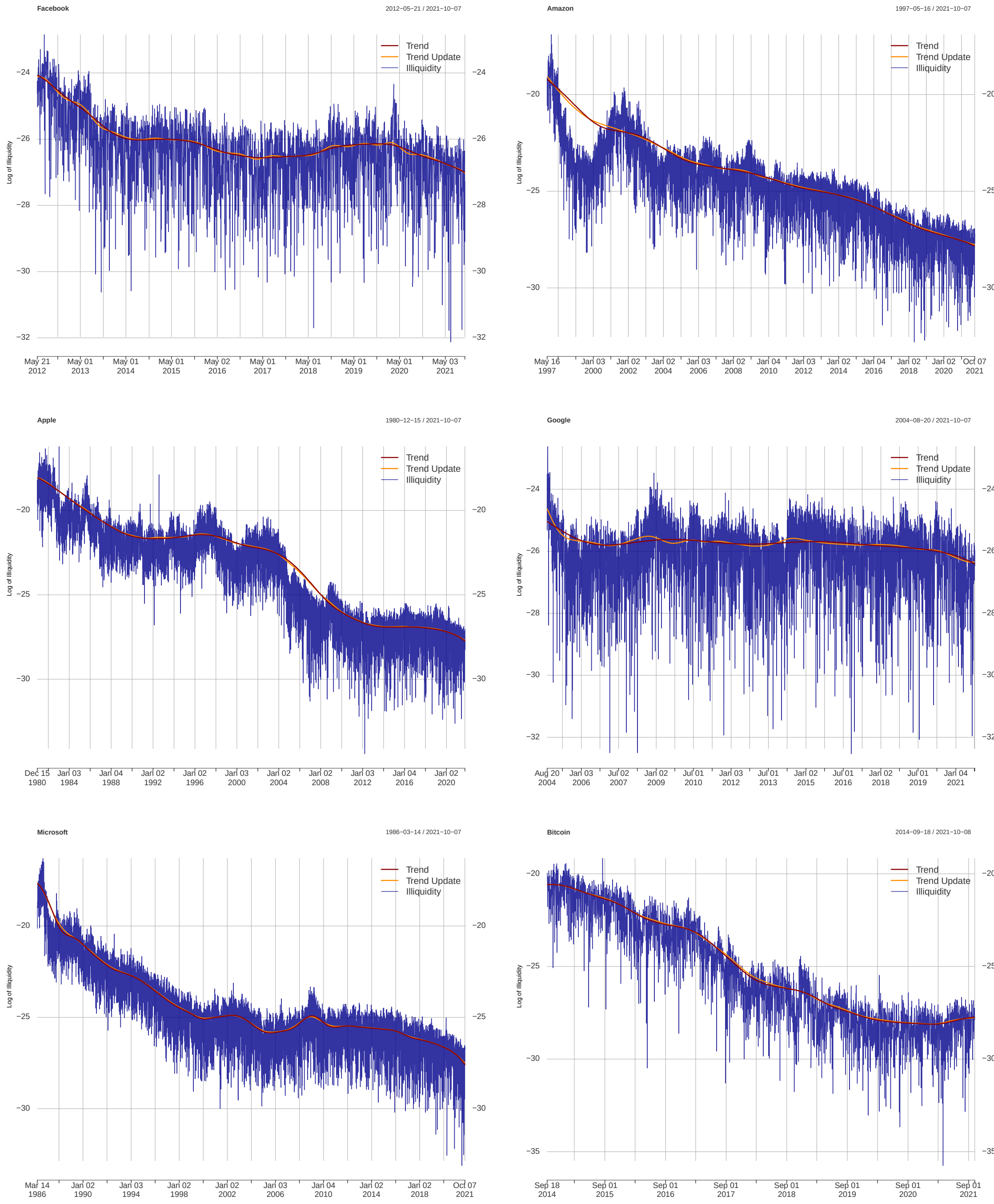


Figure 5: Fab 5 and Bitcoin log illiquidity and updated trend function based on the GMM estimator of λ_t parameters. The red curve corresponds to the initial estimate of the trend function and the yellow and green curves correspond to the updated trend estimates.

G.3 Estimation: i.i.d. error term with parametric density

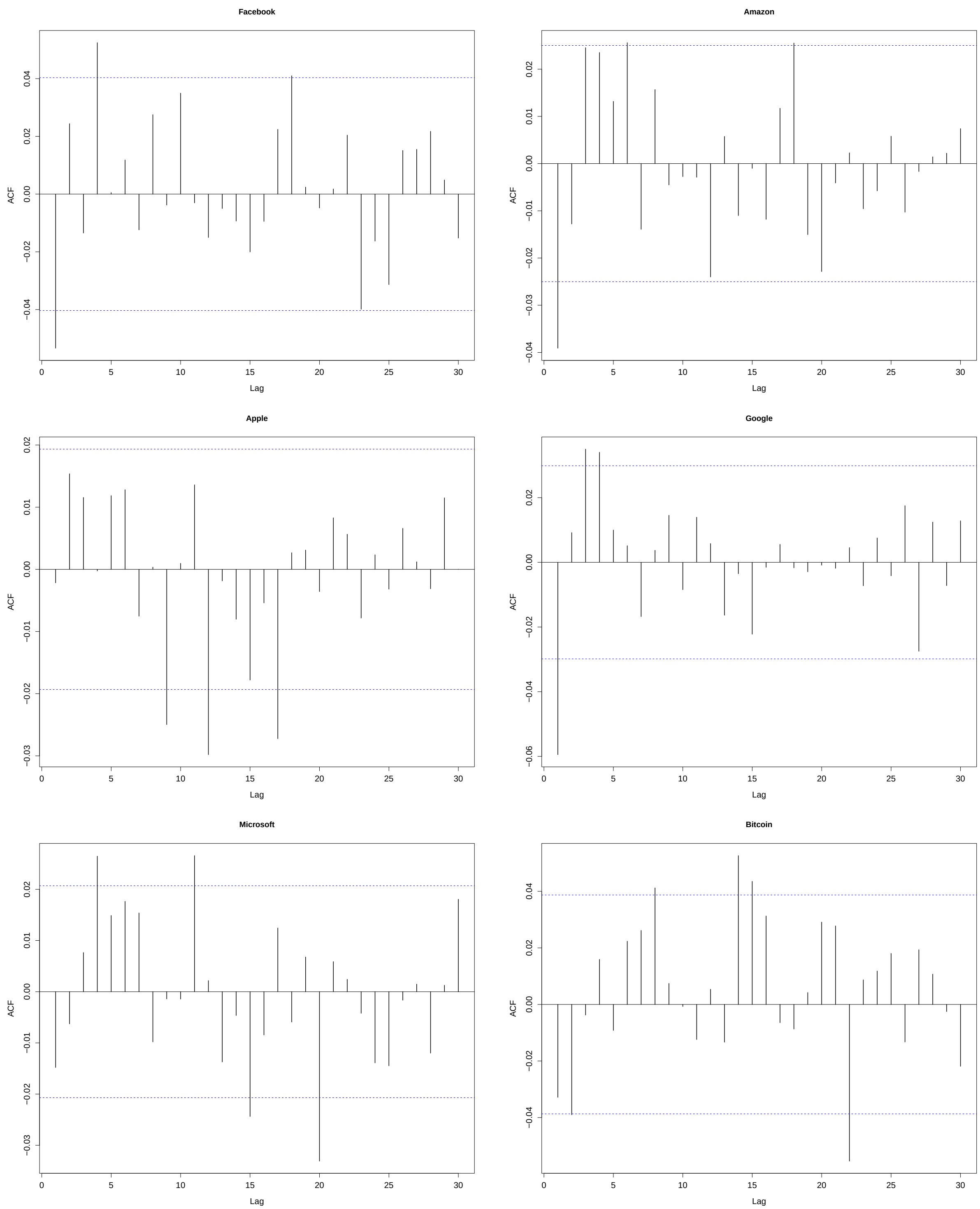


Figure 6: ACF of ζ_t.

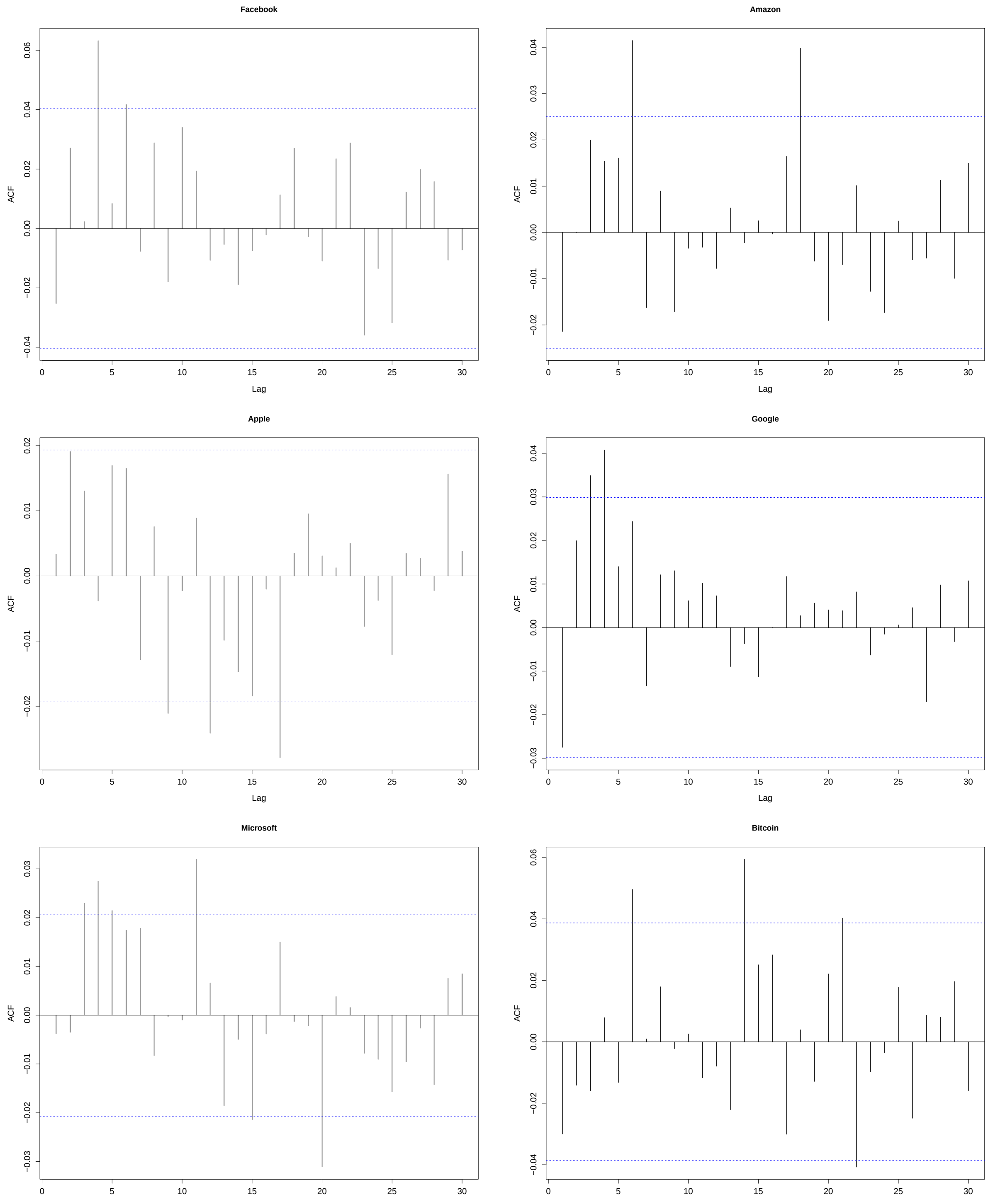


Figure 7: ACF of ζ_t^2.

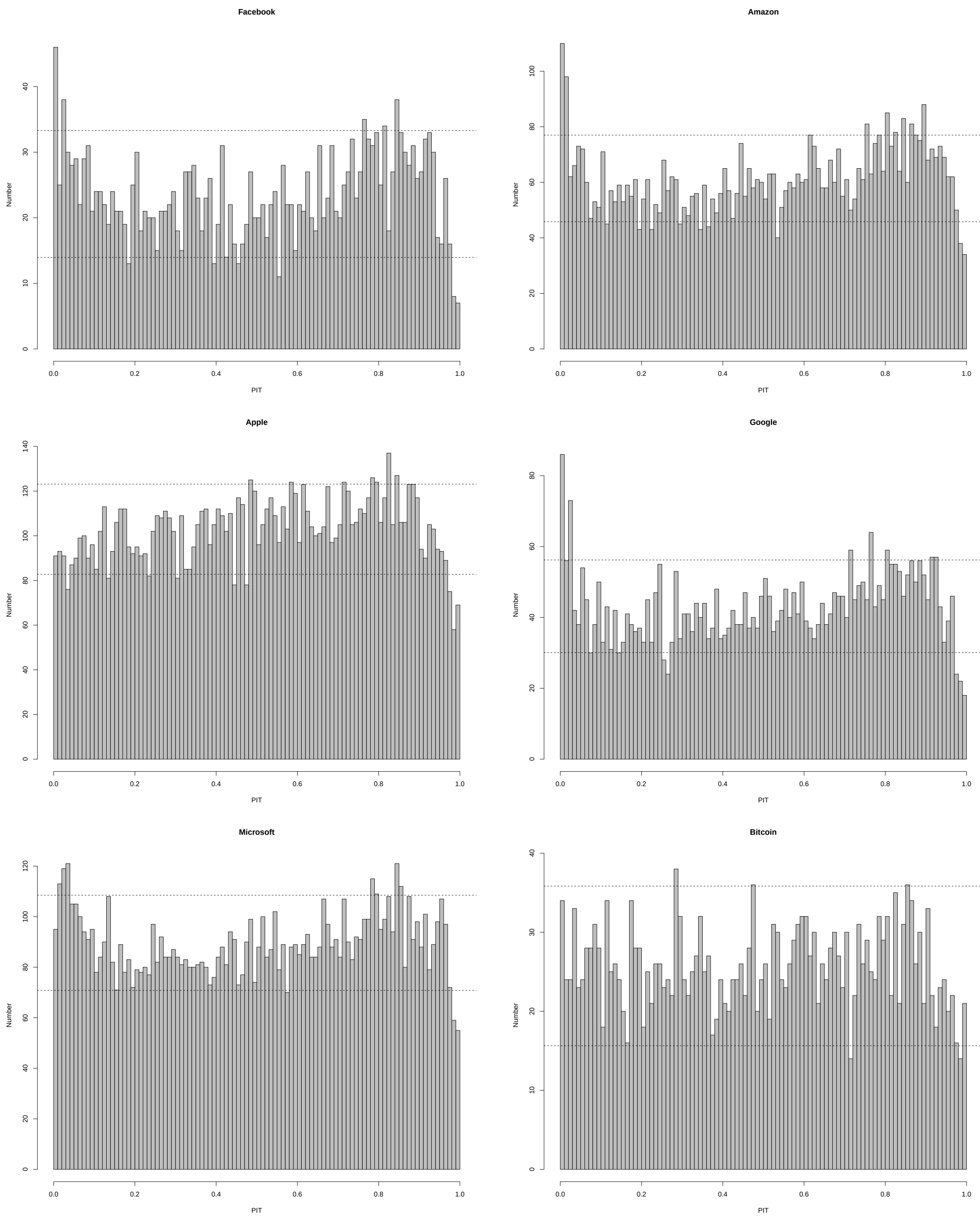


Figure 8: Probability integral transform (PIT) of ζ_t.

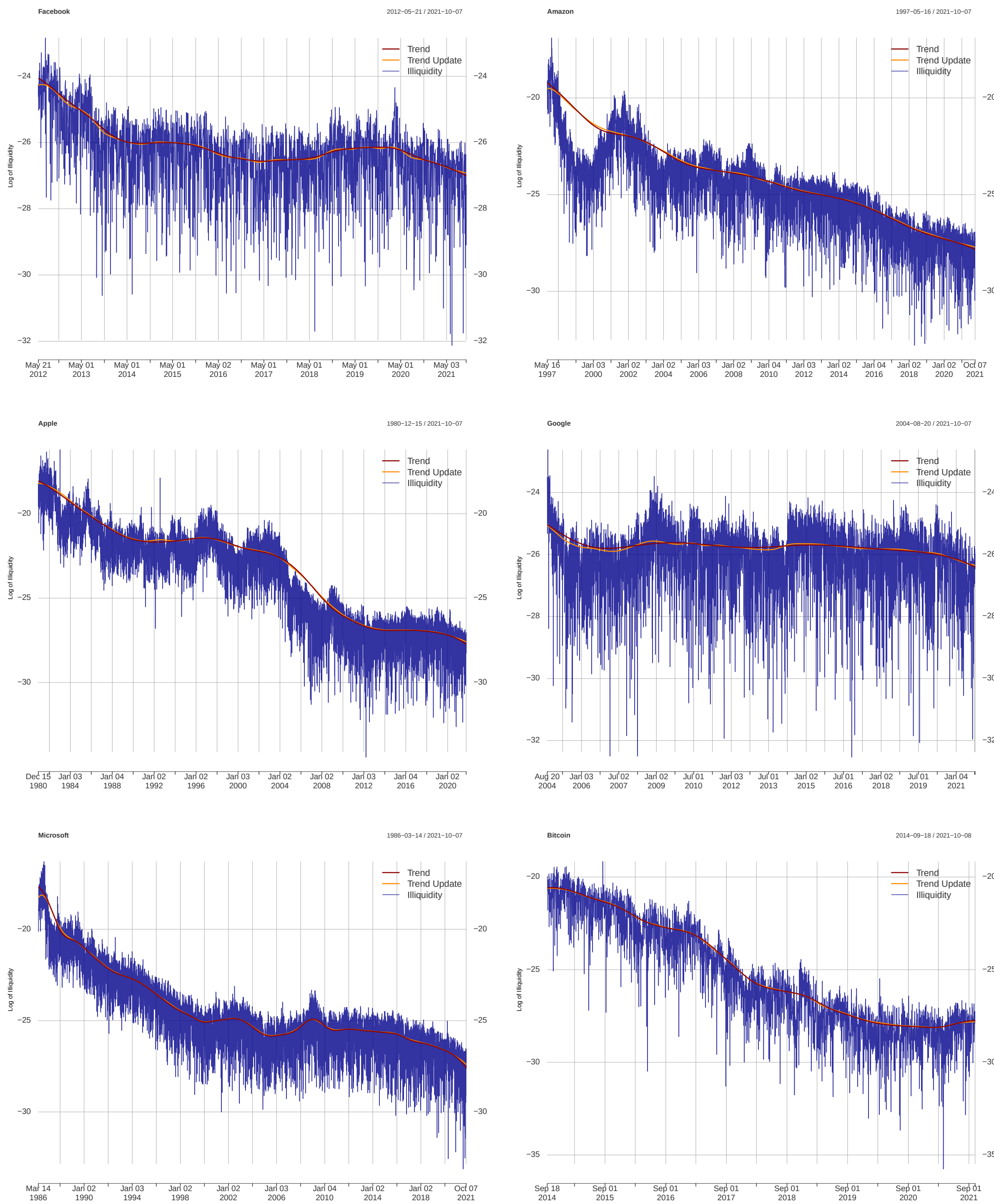


Figure 9: Fab 5 and Bitcoin log illiquidity and updated trend function based on the semiparametric ML estimator of λ_t parameters where the error term ζ_t follows a Weibull distribution. The red curve corresponds to the initial estimate of the trend function and the yellow curve corresponds to the updated trend estimates.

G.4 Estimation: i.i.d. error term with nonparametric density

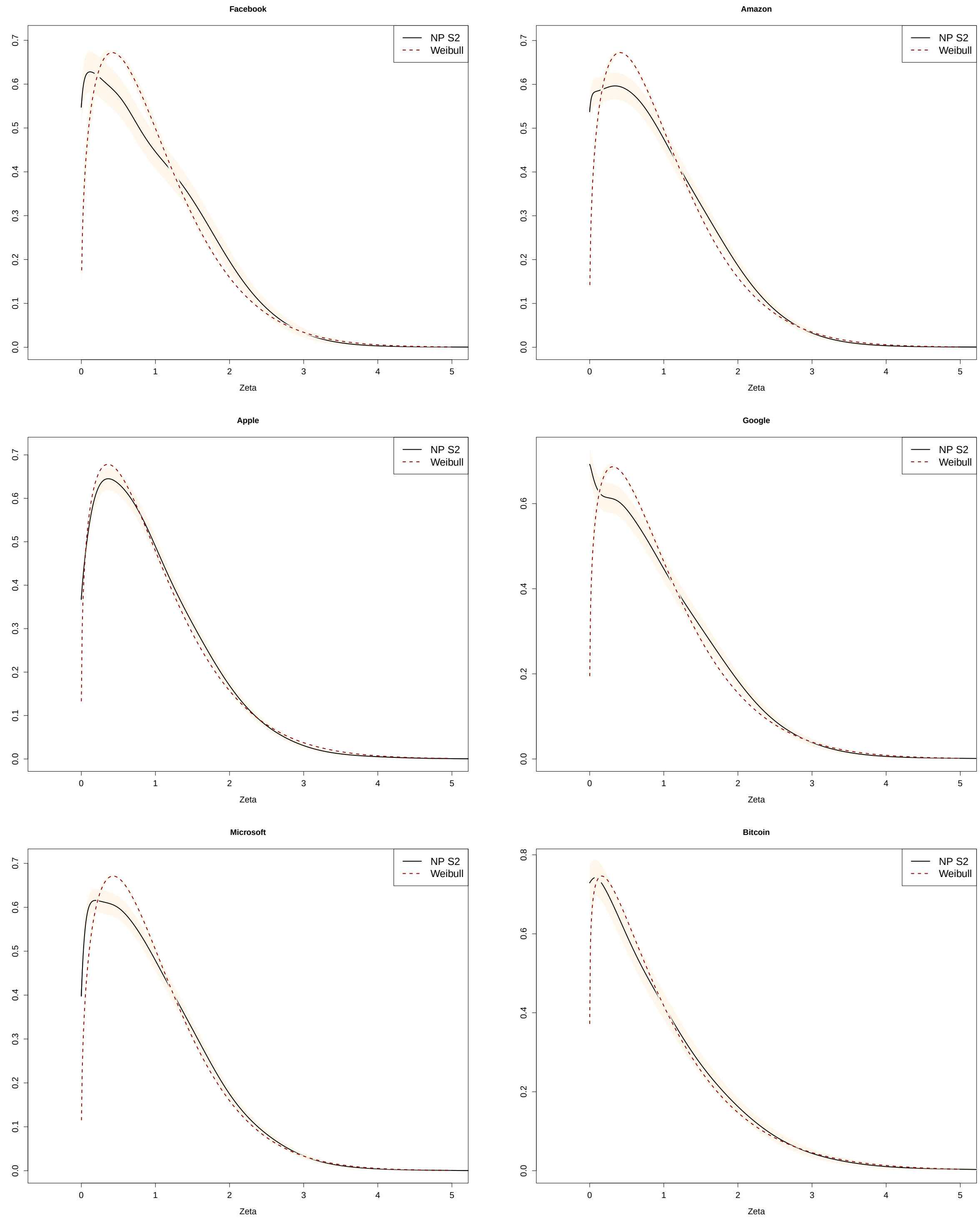


Figure 10: Comparison between the kernel density estimate of ζ_t obtained using GMM approahc (solid line) and the Weibull density (dashed line).

G.5 Risk premium

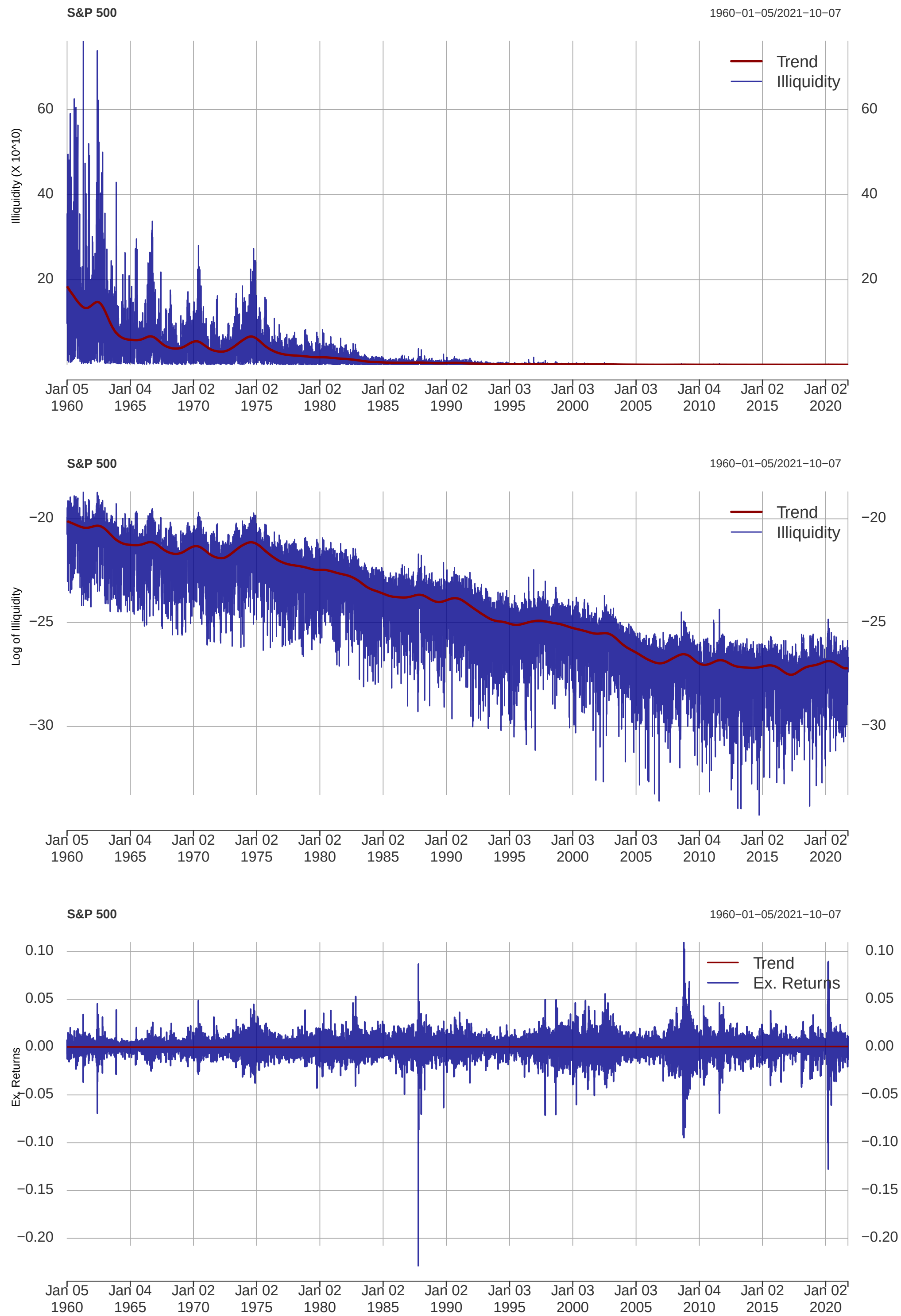


Figure 11: S&P 500 index daily (log) illiquidity series and return data.

H 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 12, 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 13 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.

We also compute the ACF of the dummy variable series and plot it in Figure 14 for 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.

[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).

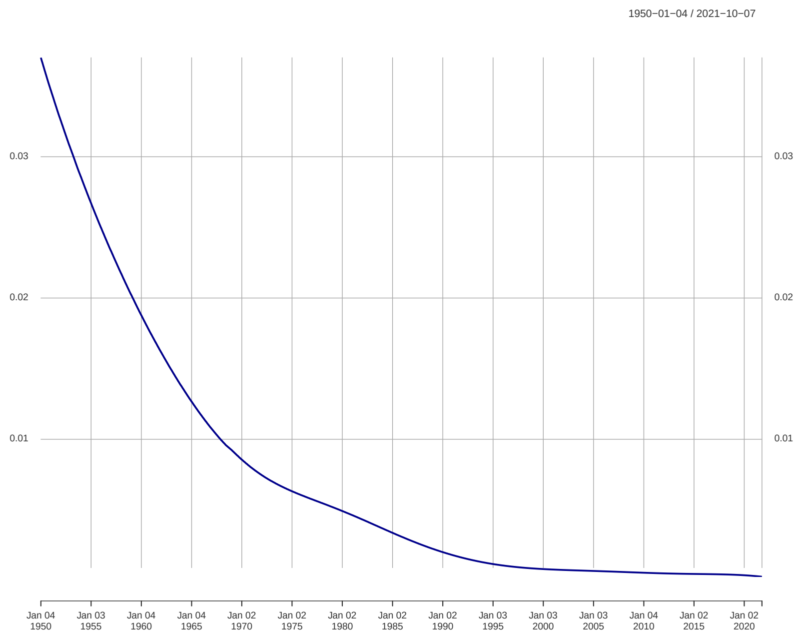


Figure 12: Smoothed series for the occurrence of exact zero returns in the S&P 500 index.

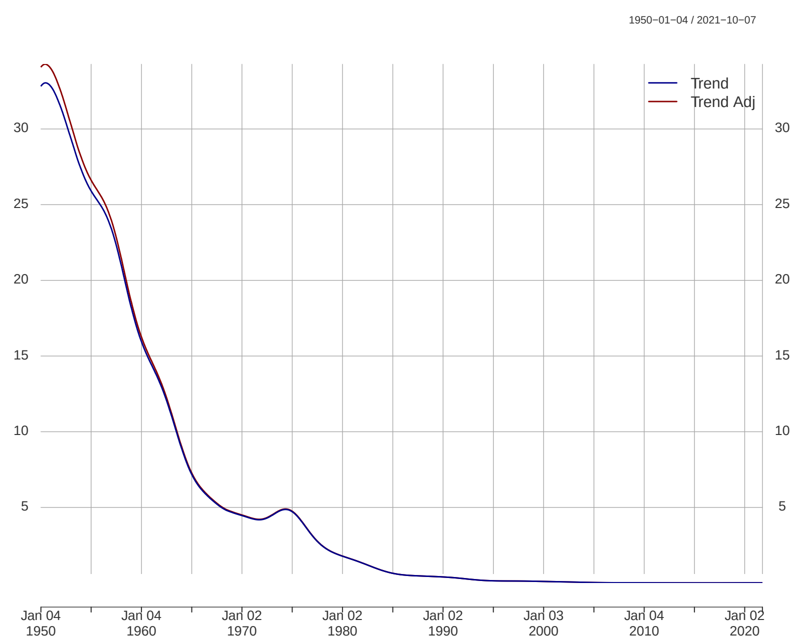


Figure 13: Illiquidity trend function $g(t/T)$ and its corresponding series adjusted for the presence of zero return observations for the S&P 500 index.

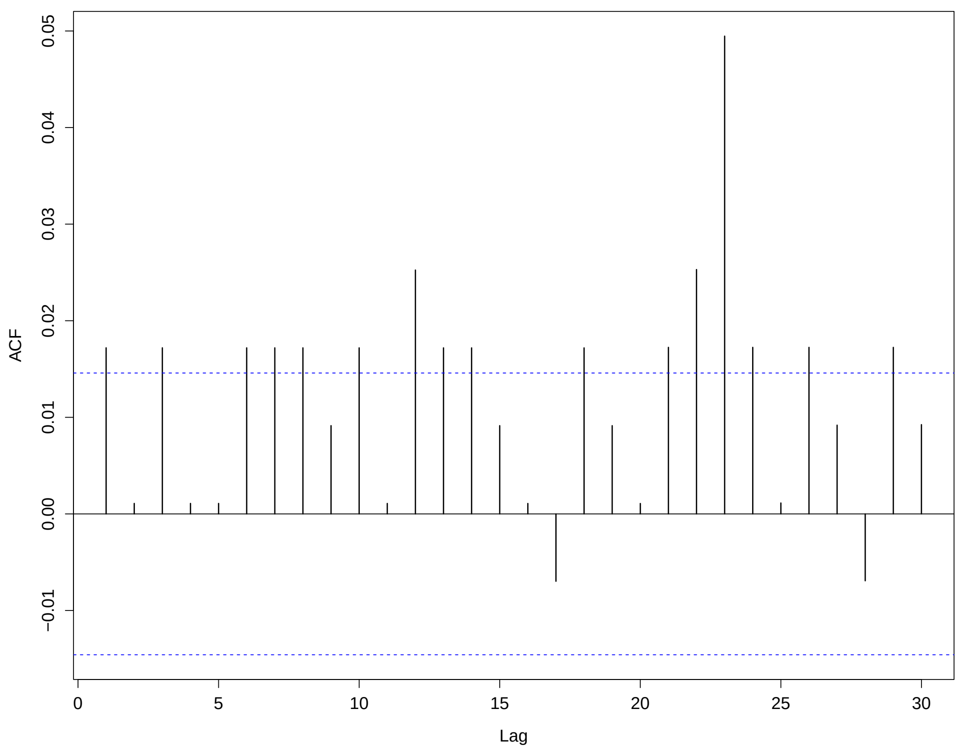


Figure 14: ACF of the series measuring the occurrence of zero return observations in the S&P 500 index.

I Tail index and fat-tailed distribution

I.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 2 the tail index estimates for the illiquidity ℓ_t, rescaled illiquidity ℓ_t^*, error term ζ_t and its reciprocal $\frac{1}{\zeta_t}$, obtained using GMM approach. The estimated tail index of rescaled illiquidity ℓ_t^* is above three for all assets. For illiquidity ℓ_t, the estimates are larger than two in most cases. The exceptions, i.e. Amazon, Apple and Microsoft, do not indicate that the first moment does not exist but rather they are due to the presence of a strong downward trend at the beginning of the sample period. This is confirmed by the estimates using only post-2000 data reported in Table 3. In this case, the estimated tail index for ℓ_t is between 1.9 and 4. In addition, 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 2: Estimated tail index.

	S&P500	Facebook	Amazon	Apple	Google	Microsoft	Bitcoin
ℓ_t	2.989 (0.14)	2.716 (0.35)	1.016 (0.08)	1.331 (0.08)	3.016 (0.29)	0.912 (0.06)	2.760 (0.35)
ℓ_t^*	4.378 (0.21)	6.278 (0.82)	3.396 (0.27)	3.480 (0.22)	4.305 (0.42)	5.632 (0.38)	4.198 (0.52)
ζ	4.784 (0.23)	7.196 (0.94)	6.428 (0.52)	3.746 (0.23)	5.502 (0.53)	5.997 (0.40)	3.370 (0.42)
$\frac{1}{\zeta}$	1.255 (0.06)	1.516 (0.20)	1.336 (0.11)	1.285 (0.08)	1.094 (0.11)	1.734 (0.12)	0.987 (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 $\log(\text{Rank} - 1/2) = a - b\log(\text{Size})$. The regression is based on the 5% of the largest observations in the distribution.

Table 3: Estimated tail index using data after 2000.

	S&P500	Facebook	Amazon	Apple	Google	Microsoft	Bitcoin
ℓ_t	3.423 (0.29)	2.716 (0.35)	1.911 (0.16)	3.282 (0.28)	3.016 (0.29)	3.986 (0.34)	2.760 (0.35)
ℓ_t^*	4.565 (0.39)	6.278 (0.82)	3.587 (0.31)	4.635 (0.40)	4.305 (0.42)	6.131 (0.52)	4.198 (0.52)
ζ	4.478 (0.38)	7.388 (0.96)	6.892 (0.59)	6.538 (0.56)	5.653 (0.55)	6.910 (0.59)	3.367 (0.42)
$\frac{1}{\zeta}$	1.026 (0.09)	1.512 (0.20)	1.339 (0.11)	1.267 (0.11)	1.098 (0.11)	1.717 (0.15)	0.987 (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 $\log(\text{Rank} - 1/2) = a - b\log(\text{Size})$. The regression is based on the 5% of the largest observations in the distribution.

I.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 4 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 177 for Apple and 3 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 0.094 for Apple, 0.052 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 7). This observation is further confirmed by Figure 15 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.

In addition, 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 likelihood ranging from 4 to 54, except in the case of Microsoft whose log likelihood under the Burr distribution is around 26 units larger. From Figure 15, 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 investigation.

Lastly, we plot, in Figure 16 to Figure 19, the estimated nonparametric density (based on the residuals $\widehat{\zeta_t}$ from the GMM approach) against the estimated Weibull, Lomax, Burr and Inverse Burr densities. We observe that there are noticeable differences between the estimated parametric and nonparametric densities, suggesting that we should further investigate whether using a nonparametric density allows us to further improve the ML estimation for the λ_t process.

Table 4: Log likelihood comparison between models using different parametric densities (Weibull, Lomax, Burr, Inverse Burr) for the error term ζ_t.

	Facebook	Amazon	Apple	Google	Microsoft	Bitcoin
Weibull	-2147.28	-3659.45	-6839.33	-3862.38	-7790.52	-2336.99
Lomax	-2299.22	-4053.17	-7449.32	-4030.13	-8425.72	-2371.10
Burr	-2147.32	-3659.39	-6662.32	-3862.39	-7790.55	-2333.91
Inv Burr	-2117.04	-3622.51	-6658.67	-3808.37	-7816.51	-2326.64

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

Table 5: Maximum likelihood estimates of the parameters for the λ_t process under the assumption that the error term ζ_t follows a Weibull distribution.

	Facebook	Amazon	Apple	Google	Microsoft	Bitcoin
β	0.886 (0.012)	0.915 (0.024)	0.901 (0.001)	0.928 (0.002)	0.917 (0.001)	0.897 (0.005)
γ	0.059 (0.015)	0.085 (0.024)	0.095 (0.000)	0.062 (0.001)	0.068 (0.001)	0.067 (0.005)
φ	1.365 (0.002)	1.355 (0.043)	1.307 (0.001)	1.266 (0.000)	1.379 (0.002)	1.135 (0.023)
σ_ζ	0.741	0.746	0.772	0.795	0.734	0.883

Note: The estimated parameters are $\theta = (\beta, \gamma, \varphi)$ for the λ_t process. 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 standard errors of the corresponding parameter estimates. We note that the standard errors are underestimated as they do not account for the estimation error associated with the smoothed liquidity process in the first step of the estimation.

Table 6: Maximum likelihood estimates of the parameters for the λ_t process under the assumption that the error term ζ_t follows a Lomax distribution.

	Facebook	Amazon	Apple	Google	Microsoft	Bitcoin
β	0.878 (0.026)	0.915 (0.007)	0.906 (0.007)	0.929 (0.011)	0.917 (0.001)	0.899 (0.018)
γ	0.060 (0.013)	0.085 (0.007)	0.090 (0.007)	0.060 (0.010)	0.069 (0.001)	0.068 (0.011)
α	307.686 (42.008)	224.827 (0.780)	140.772 (1.954)	256.016 (20.211)	991.652 (0.001)	100.071 (0.805)
σ_ζ	1.003	1.004	1.007	1.004	1.001	1.010

Note: The estimated parameters are $\theta = (\beta, \gamma, \alpha)$ for the λ_t process. 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 standard errors of the corresponding parameter estimates. We note that the standard errors are underestimated 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 Burr distribution.

	Facebook	Amazon	Apple	Google	Microsoft	Bitcoin
β	0.889 (0.000)	0.915 (0.000)	0.909 (0.004)	0.927 (0.000)	0.915 (0.000)	0.900 (0.005)
γ	0.058 (0.000)	0.085 (0.000)	0.088 (0.004)	0.062 (0.000)	0.069 (0.000)	0.069 (0.002)
γ^B	1.361 (0.000)	1.355 (0.000)	1.462 (0.001)	1.267 (0.000)	1.381 (0.000)	1.175 (0.005)
λ^B	0.000 (0.000)	0.000 (0.000)	0.094 (0.000)	0.000 (0.000)	0.000 (0.000)	0.052 (0.004)
σ_ζ	0.743	0.746	0.748	0.795	0.733	0.895

Note: The estimated parameters are $\theta = (\beta, \gamma, \gamma^B, \lambda^B)$ for the λ_t process. (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 standard errors of the corresponding parameter estimates. We note that the standard errors are underestimated 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 an Inverse Burr distribution.

	Facebook	Amazon	Apple	Google	Microsoft	Bitcoin
β	0.886 (0.016)	0.917 (0.006)	0.908 (0.006)	0.930 (0.005)	0.919 (0.005)	0.907 (0.007)
γ	0.060 (0.008)	0.083 (0.006)	0.088 (0.006)	0.059 (0.004)	0.068 (0.004)	0.067 (0.006)
τ	6.320 (0.028)	5.178 (0.024)	4.139 (0.022)	4.871 (0.002)	4.483 (0.023)	3.511 (0.032)
α	0.146 (0.003)	0.185 (0.004)	0.268 (0.005)	0.185 (0.004)	0.233 (0.003)	0.259 (0.008)
σ_ζ	0.698	0.731	0.765	0.773	0.749	0.938

Note: The estimated parameters are $\theta = (\beta, \gamma, \tau, \alpha)$ for the λ_t process. (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 standard errors of the corresponding parameter estimates. We note that the standard errors are underestimated as they do not account for the estimation error associated with the smoothed liquidity process in the first step of the estimation.

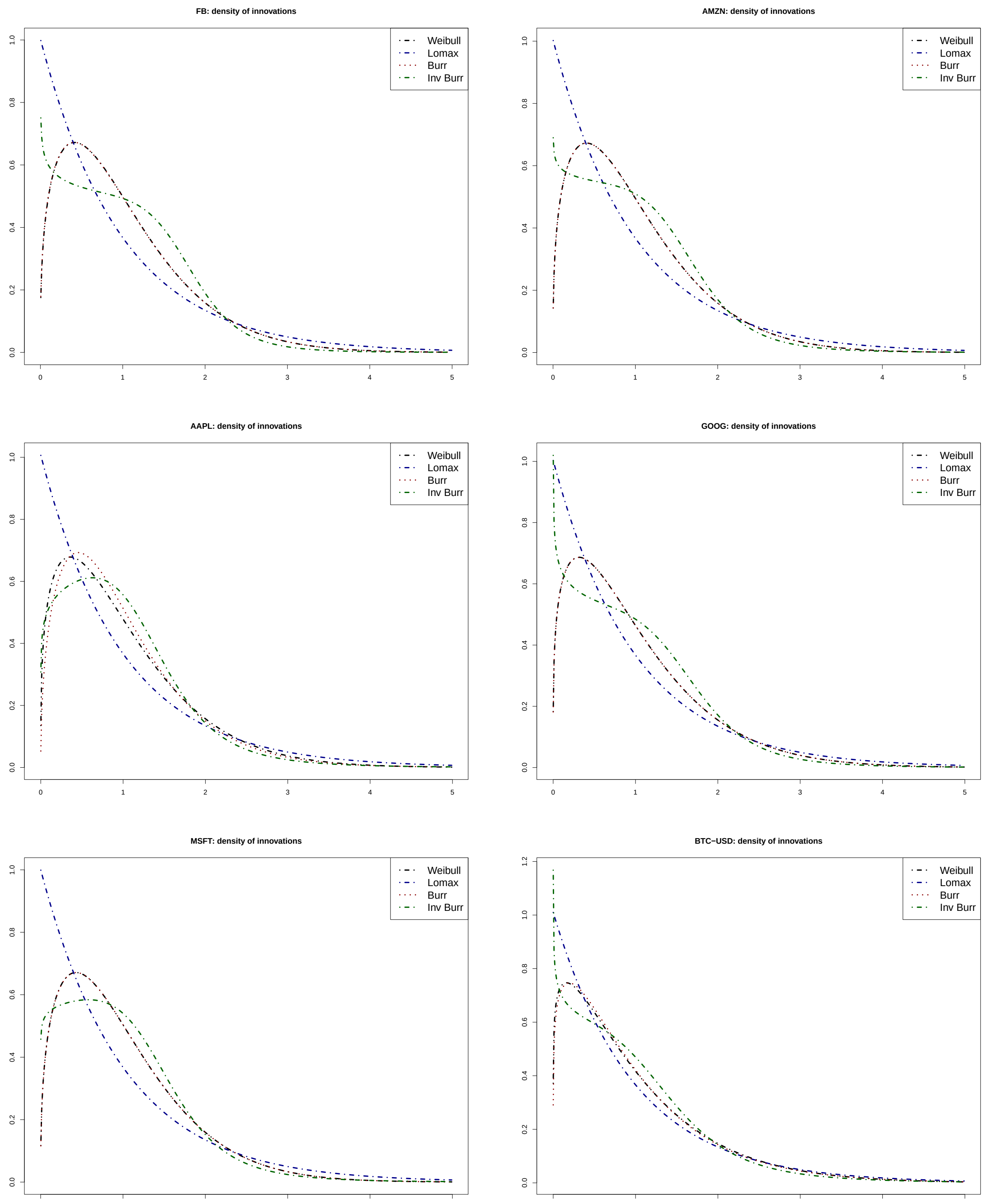


Figure 15: Comparison between the estimated Weibull, Lomax, Burr and Inverse Burr densities.

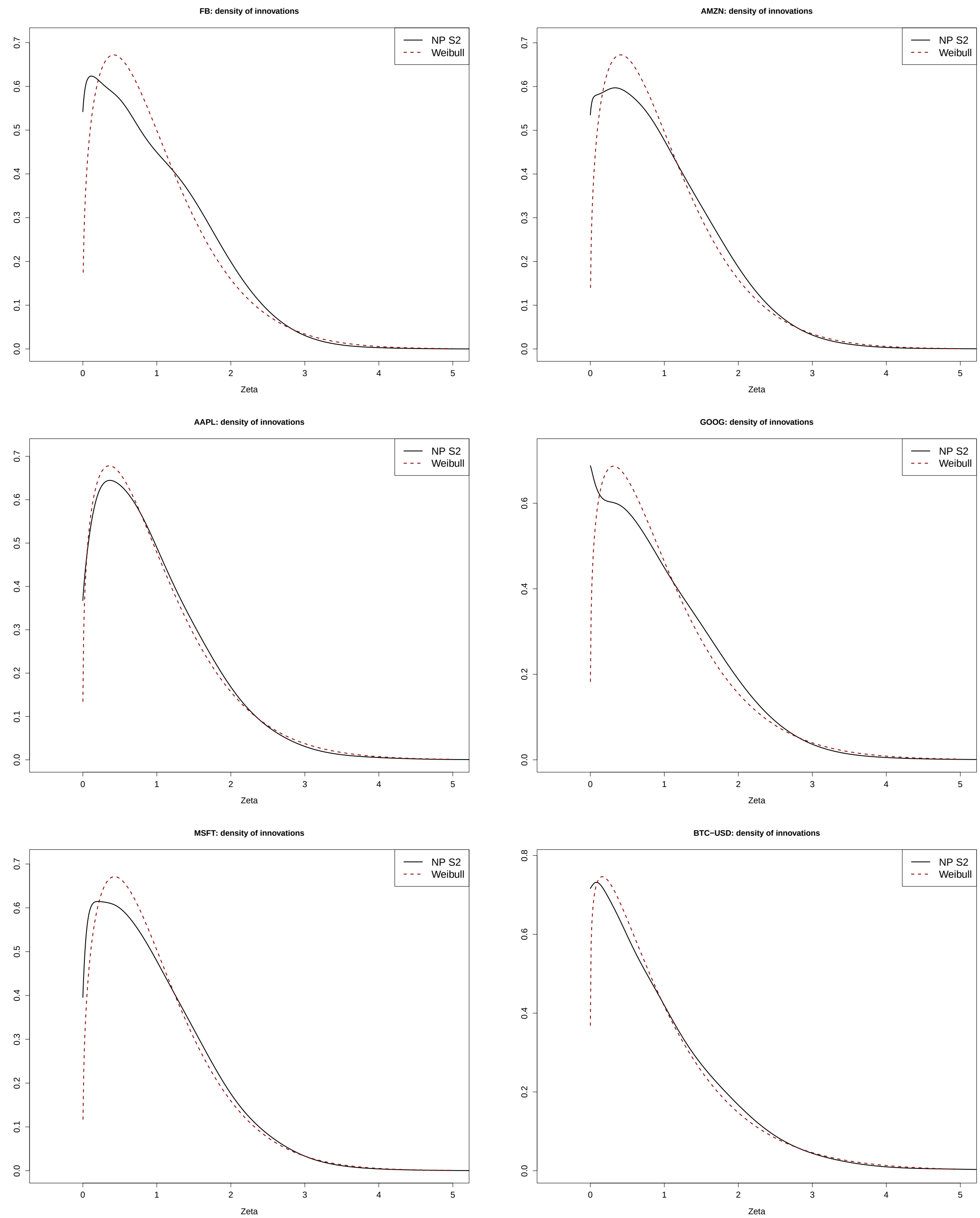


Figure 16: Comparison between the kernel density estimate of ζ_t (solid line) and the Weibull density (dashed line).

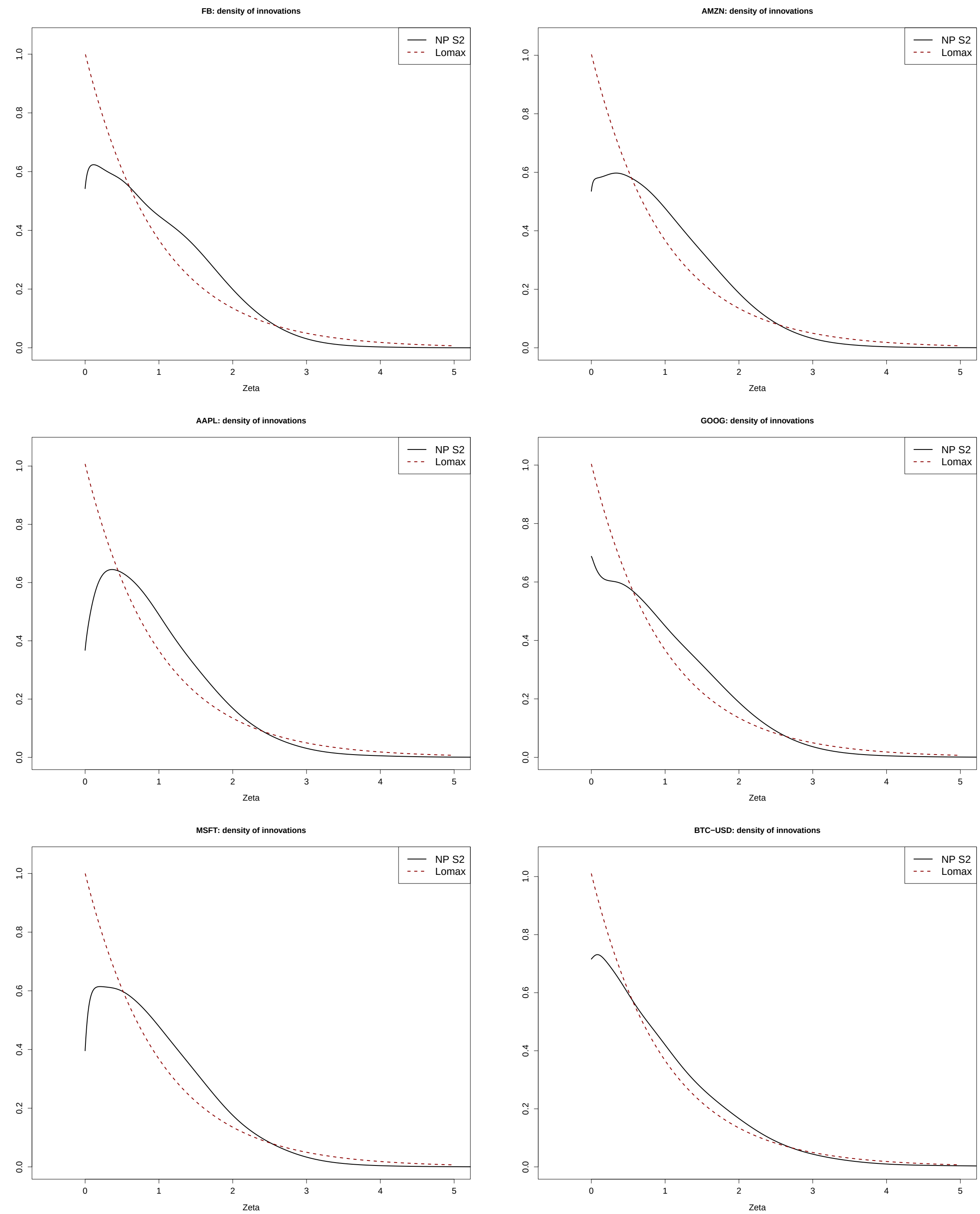


Figure 17: Comparison between the kernel density estimate of ζ_t (solid line) and the Lomax density (dashed line).

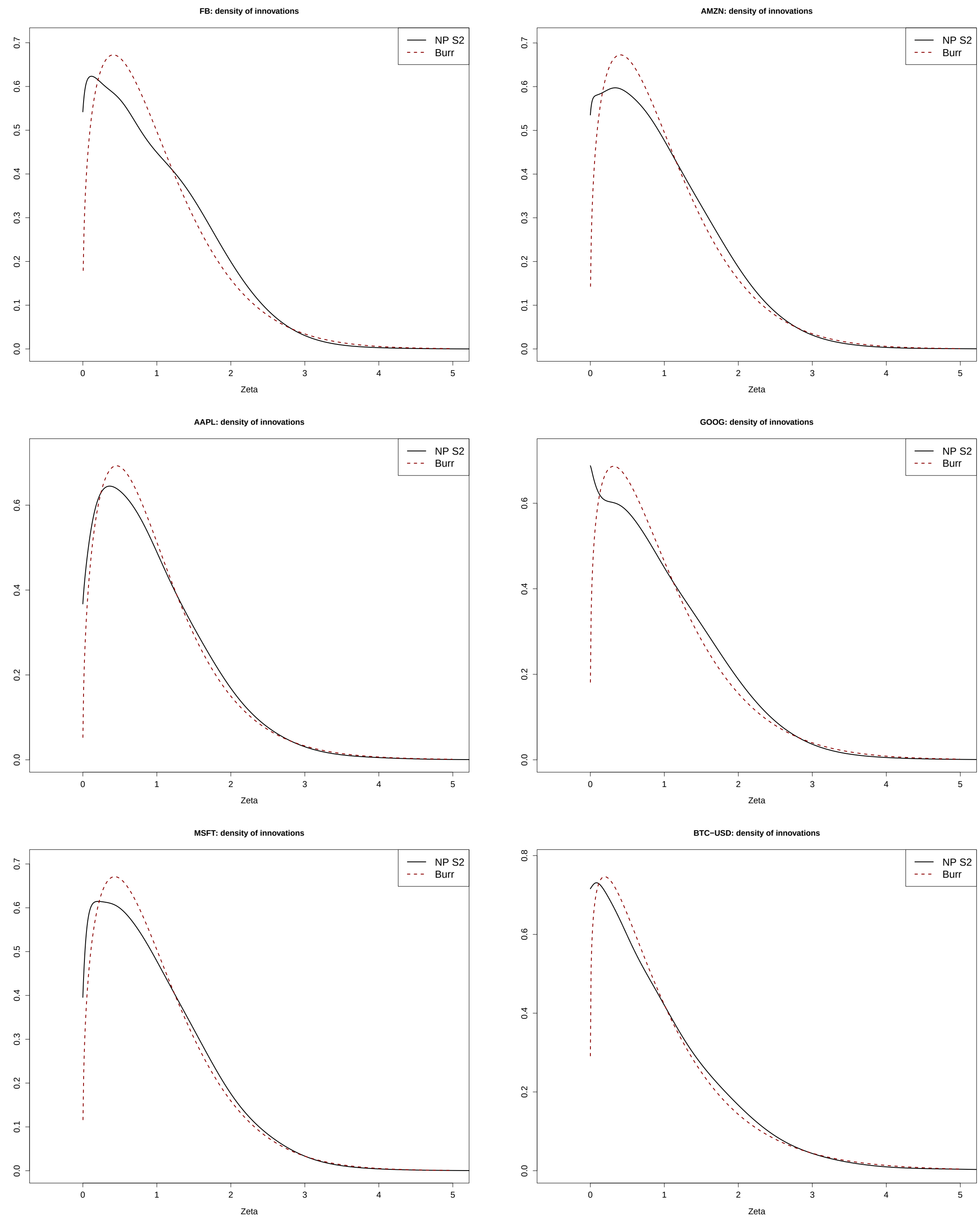


Figure 18: Comparison between the kernel density estimate of ζ_t (solid line) and the Burr density (dashed line).

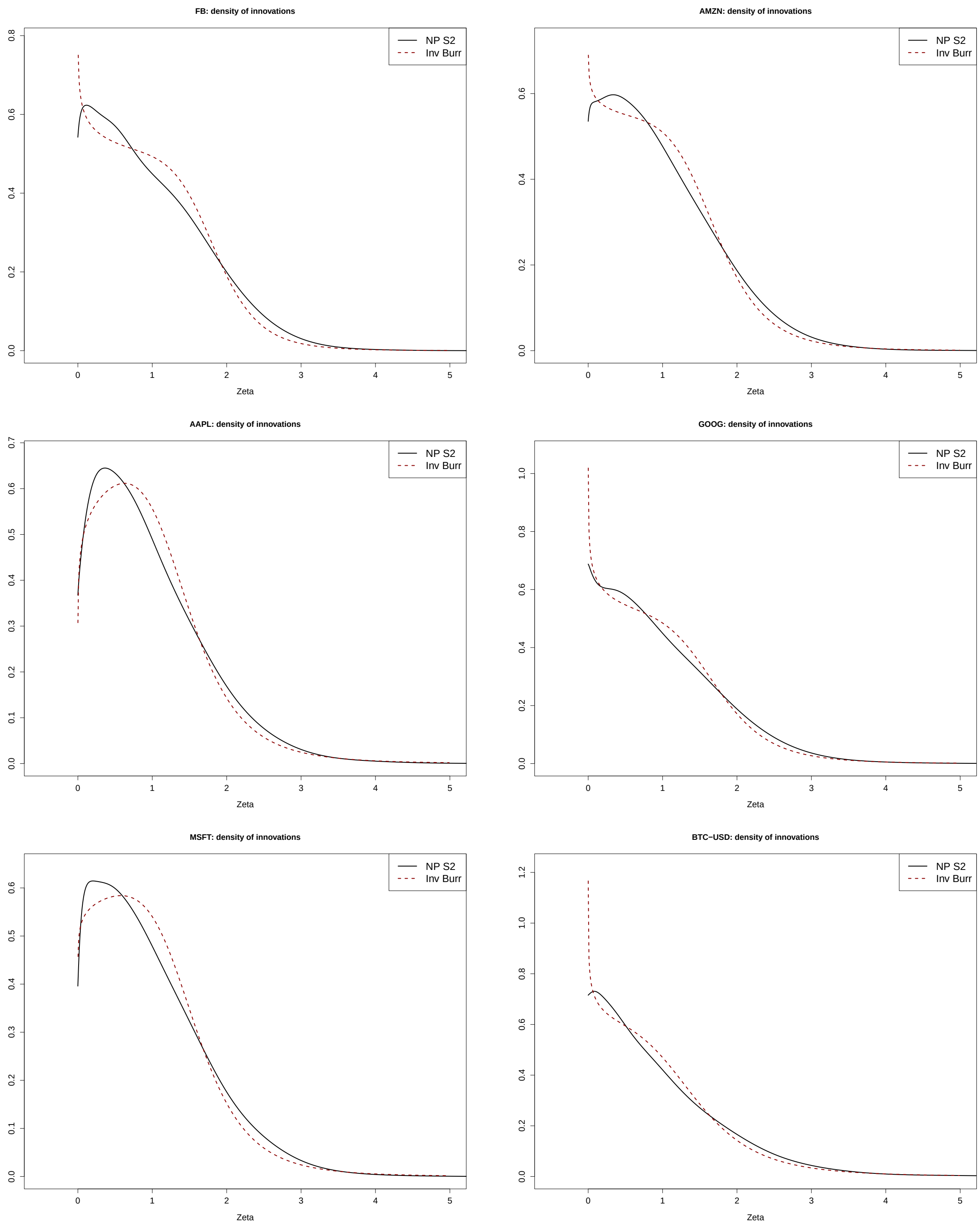


Figure 19: Comparison between the kernel density estimate of ζ_t (solid line) and the Inverse Burr density (dashed line).

J Mean smoothing and median smoothing

In our model, we assume that the illiquidity process follows a multiplicative process as

$$\ell_t = g(t/T)\lambda_t\zeta_t \tag{17}$$

$$\lambda_t = \omega + \beta\lambda_{t-1} + \gamma\ell^*_{t-1}, \tag{18}$$

where $g(.)$ is a positive and smooth but unknown function of rescaled time, $\ell^*_t = \ell_t/g(t/T)$ is the rescaled illiquidity, and ζ_t is a non-negative random variable with conditional mean one and finite unconditional variance.

We first estimate the trend function $g(.)$ using the following estimator

- Mean smoothing: first, we have

$$(\widehat{\alpha}, \widehat{\beta}) = \arg\min_{\alpha,\beta} \sum_{t=1}^{T} K_h(t/T-u)\left\{\ell_t - \alpha - \beta(t/T-u)\right\}^2,$$

- Median smoothing: compute the 3-day rolling window median (ℓ_t^{med}) of the illiquidity series, then we have

$$(\widehat{\alpha}, \widehat{\beta}) = \arg\min_{\alpha,\beta} \sum_{t=1}^{T} K_h(t/T-u)\left|\ell_t^{\text{med}} - \alpha - \beta(t/T-u)\right|,$$

and let $\widehat{g}^*(u) = \widehat{\alpha}$ for each $u \in [0,1]$. In both cases, we normalize them such that they integrate to one, i.e. let

$$\widehat{g}(u) = \frac{\widehat{g}^*(u)}{\int_0^1 \widehat{g}^*(u)du}.$$

The observed illiquidity series together with the estimated trend functions, using mean smoothing and median smoothing methods, are plotted in Figure 20. The normalized trend functions are plotted in Figure 21.

We then obtain the detrended illiquidity series and estimate the dynamic parameters of the λ_t process using a one-step GMM approach. The parameter estimates and their associated standard errors are reported in Table 9.

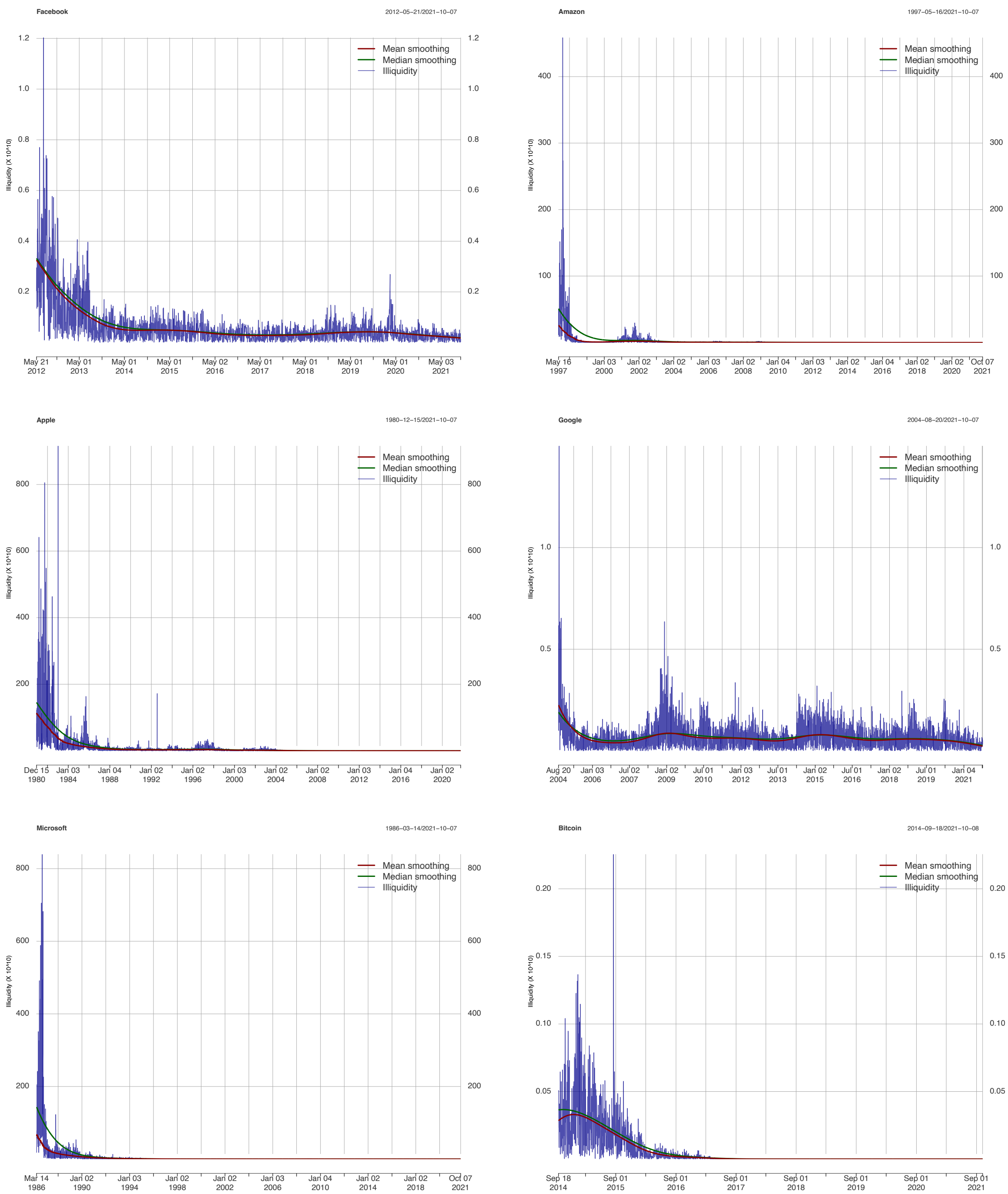


Figure 20: Fab 5 and Bitcoin daily illiquidity series and its trend.

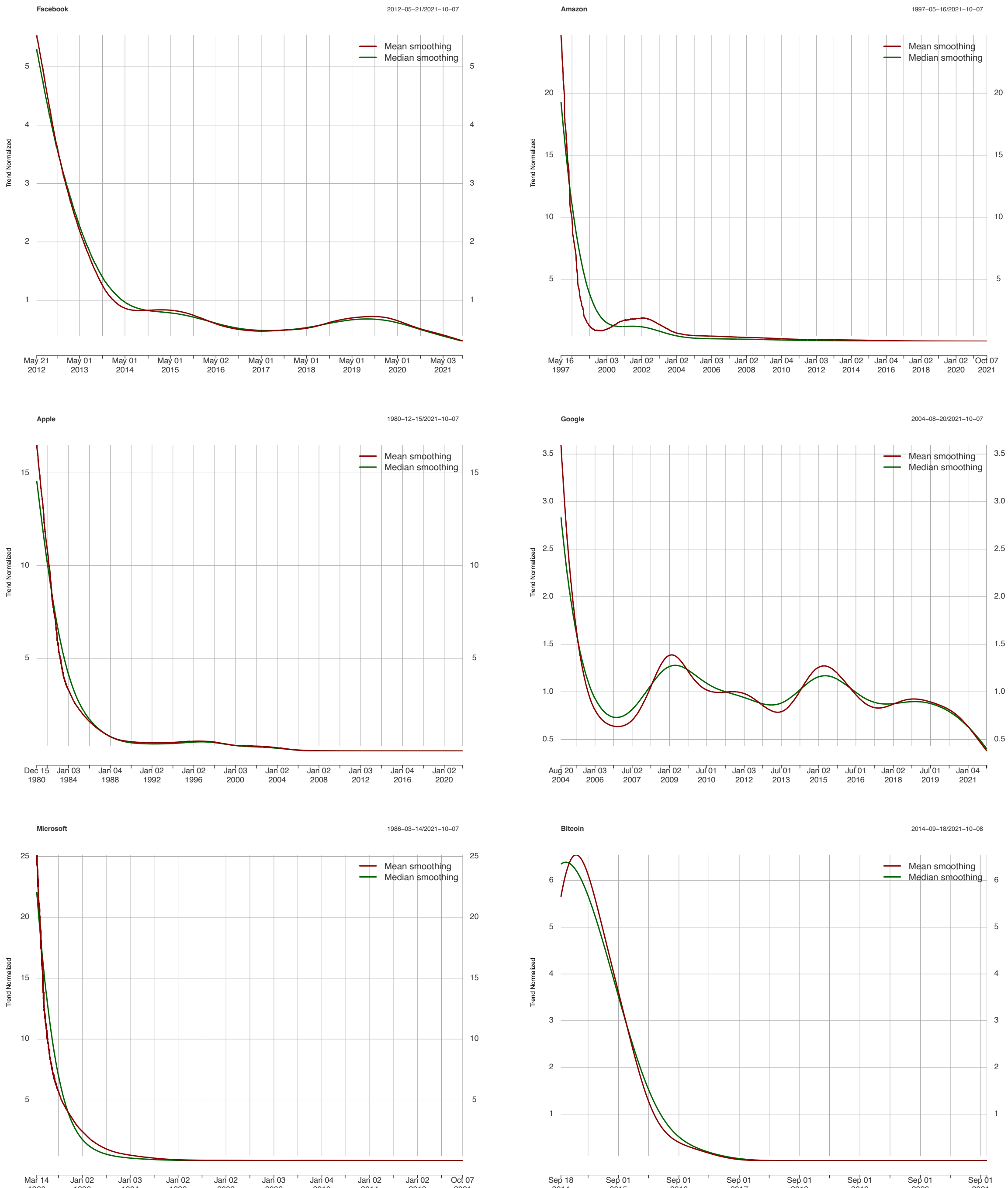


Figure 21: Normalized illiquidity trend for Fab 5 and Bitcoin.

Table 9: Estimated parameters of the λ_t process based on first moment restriction.

	Mean smoothing			Median smoothing		
	ω	β	γ	ω	β	γ
Facebook	0.006	0.956	0.034	0.008	0.960	0.027
	(0.005)	(0.018)	(0.013)	(0.006)	(0.021)	(0.013)
Amazon	0.009	0.943	0.053	0.013	0.940	0.053
	(0.007)	(0.016)	(0.013)	(0.007)	(0.013)	(0.010)
Apple	0.014	0.899	0.085	0.014	0.900	0.083
	(0.004)	(0.015)	(0.011)	(0.004)	(0.014)	(0.010)
Google	0.007	0.960	0.030	0.005	0.965	0.027
	(0.004)	(0.014)	(0.009)	(0.004)	(0.014)	(0.009)
Microsoft	0.007	0.932	0.057	0.005	0.936	0.054
	(0.003)	(0.012)	(0.008)	(0.002)	(0.010)	(0.007)
Bitcoin	0.018	0.895	0.067	0.031	0.882	0.058
	(0.008)	(0.025)	(0.013)	(0.019)	(0.042)	(0.012)

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. (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.

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

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.

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.

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

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

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

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

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.