Supplementary Material: Dependence Studies

Our formal analysis is based on the hypothesis that the size data are i.i.d. Results in Hsing (1991) suggest that the current theory might be a reasonable guide if the dependence is sufficiently weak so that approximations to a normal law still hold. In this appendix this conjecture is verified for dependent data generating processes also considered in Gabaix and Ibragimov (2011).

1 AR(1) and MA(1) Processes

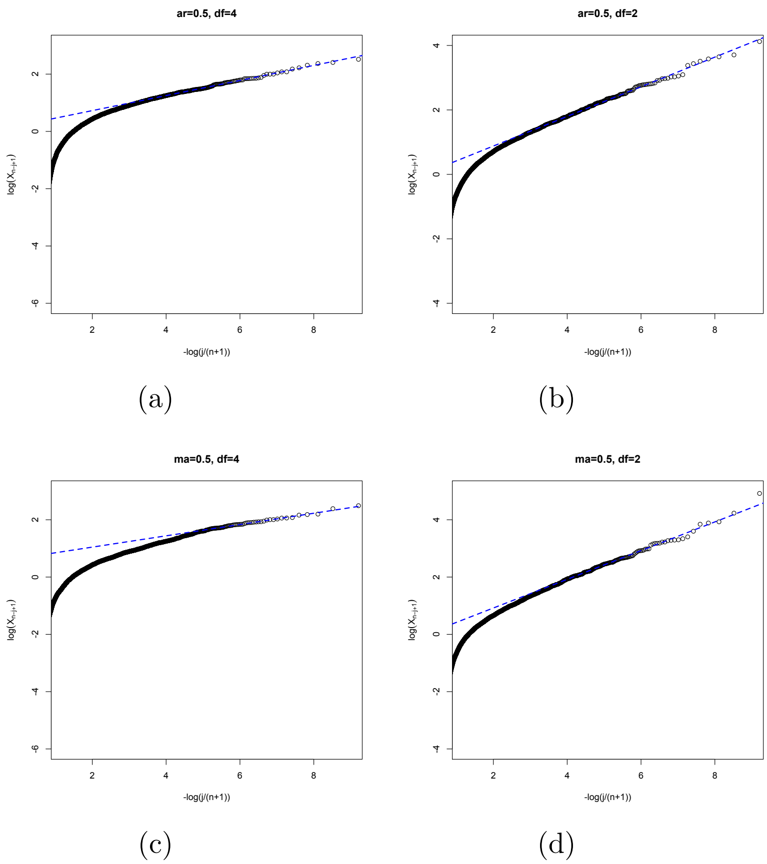
In particular, we model dependence by AR(1) and MA(1) processes. In order to study the performance of the regression estimator $\hat{\gamma}$, we assume that heavy-tailed innovations follow a student-t distribution with *df* degrees of freedom with $df \in \{2, 4\}$. Recall that $\gamma = 1/df$.

In order to illustrate the data and the performance of our methods, we depict in Figure 1 Pareto QQ-plots for single random samples of size $N = 10,000$ of the DGP when the auto-regressive and moving-average parameter equals 0.5 and the student-t distributed innovation has *df* degrees of freedom. Also depicted is the Zipf-regression with slope $\hat{\gamma}(k^*)$ and threshold $X_{n-k^*,n}$ where our selection procedure determines k^*. It is evident that in all cases depicted, the Pareto QQ-plot has a concave-like shape, and that our method selects well the ultimate linear part of the QQ-plot. If k is to be selected arbitrarily at, for instance, 5% of the sample, so $k = n_1 = 500$, the plots also suggest that $\hat{\gamma}(n_1)$ should perform reasonably well, as this corresponds to an abscissa value of about 3, except perhaps for the MA(1) model where $df = 4$ (Figure 1.c) since at this value the curvature of the plot is already pronounced. This is confirmed in the Monte Carlo below, where the mean of $\hat{\gamma}(500)$ equals 0.310 (with sd .015);[1] a 95% symmetric confidence interval using this value would exclude the

[1] The standard deviation sd is calculated across all iterations of the Monte Carlo, and closely matches the prediction based on the asymptotic theory. In particular, the standard error of $\hat{\gamma}(500)$ is $\sqrt{5/4} * .31/\sqrt{500} = .0155$. Hence, for the sake of brevity, only sd is reported below.

population value. By contrast, our selection procedure yields mean values of k^* of 244 and of $\hat{\gamma}(\bar{k}^*)$ of .285 (with sd=.021).

Figure 1: Pareto QQ-plots for AR(1) and MA(1) models, and the Zipf-regression estimate.



Notes. Sample size $N = 10,000$. Dashed lines: Zipf-regression with slope $\hat{\gamma}(k^*)$ and threshold $X_{n-k^*,n}$ where our selection procedure determines k^*. Panels (a) and (b): One random sample of a AR(1) process with auto-regressive parameter 0.5. Panels (c) and (d): One random sample of a MA(1) process with moving-average parameter 0.5. $\gamma = .25$ in panels (a) and (c), whilst $\gamma = .5$ in panels (b) and (d).

We proceed to examine the performance of the Zipf regression in Monte Carlo studies, drawing samples of size $N = 10,000$ and using $R = 1,000$ repetitions, varying the parameters of the AR(1) and MA(1) process, as well as the degrees of freedom of the student-t distribution of innovations. Tables 1 and 2 report the results. n_1 is set mechanically at 5% of N, and we report mean values of $\hat{\gamma}(n_1)$ and $\hat{\gamma}(\bar{k}^*)$.

Table 1: Performance evidence: AR(1) models.

AR	$\gamma = 0.25$				$\gamma = 0.5$			
	n_1	mean $\hat{\gamma}(n_1)$	$\bar{k}^*$	mean $\hat{\gamma}(\bar{k}^*)$	n_1	mean $\hat{\gamma}(n_1)$	$\bar{k}^*$	mean $\hat{\gamma}(\bar{k}^*)$
0	500	0.334	230	0.305	500	0.537	374	0.529
		(.014)	(119)	(.020)		(.025)	(132)	(.029)
0.25	500	0.325	237	0.298	500	0.523	388	0.518
		(.0135)	(120)	(.019)		(.028)	(125)	(.031)
0.5	500	0.302	237	0.275	500	0.501	405	0.499
		(.015)	(119)	(.021)		(.035)	(119)	(.038)
0.75	500	0.267	214	0.224	500	0.469	348	0.466
		(.0188)	(126)	(.027)		(.052)	(161)	(.063)

Notes. Monte Carlo with samples of size $N = 10,000$ and $R = 1,000$ repetitions. n_1 is a fixed cut-off set at 5% of the sample size. $\bar{k}^*$ is the mean of k^*. The DGP is an AR(1) process with stated auto-regressive parameter and innovations drawn from a student-t distribution with $1/\gamma$ degrees of freedom. Standard deviations for each experiment in parenthesis.

Before examining the dependent processes, it is of interest to examine the non-dependent process, i.e. the distribution of innovations. The results are reported in row 1 of Table 1. The relative distortions are larger when $df = 4$ compared to $df = 2$, which is expected since then $\rho = -2/df$, and the nuisance function in the distributional model given by equation (1) in the paper declines more slowly. Consider first the case of $df = 4$, so $\gamma = .25$. $\bar{k}^*$ equals about a half of n_1 and results in a lower mean point estimate. However, both $\hat{\gamma}(n_1)$ and $\hat{\gamma}(\bar{k}^*)$ exceed the population value. Since the student-t distribution is a member of the Hall class, we can use the stated asymptotic theory to compute the theoretical distortion. Using equation (5) of the paper with $A(t) = (\rho^2/\gamma)dt^{\rho}$, $\rho = -2/df$ and d given explicitly in Appendix A.3, we calculate $b_{230,1000}$ to equal .04, the bias corrected $\hat{\gamma}(230)$, being 0.26, is now close to the population value. Similar observation apply to the case $df = 2$, the theoretical bias correction at 374 evaluating to 0.02.

We turn to the dependent processes. Throughout all models, the mean point estimates decline as dependence (as measured by the first auto-correlation) increases.

Table 2: Performance evidence: MA(1) models.

MA	$\gamma = 0.25$				$\gamma = 0.5$			
	n_1	mean $\hat{\gamma}(n_1)$	$\bar{k}^*$	mean $\hat{\gamma}(\bar{k}^*)$	n_1	mean $\hat{\gamma}(n_1)$	$\bar{k}^*$	mean $\hat{\gamma}(\bar{k}^*)$
0.25	500	0.325	243	0.299	500	0.525	390	0.520
		(.014)	(124)	(.019)		(.029)	(126)	(.033)
0.5	500	0.310	244	0.285	500	0.512	388	0.508
		(.015)	(123)	(.021)		(.032)	(129)	(.036)
0.75	500	0.300	228	0.271	500	0.506	380	0.502
		(.016)	(124)	(.024)		(.032)	(140)	(.037)

Notes. As per Table 1. Here the DGP is a MA(1) process with stated moving-average parameter.

Based on the preceding exploratory analysis depicted in Figure 1, we expect $\hat{\gamma}(\bar{k}^*)$ to perform well. When df=4 we observe that $\bar{k}^*$ equals about a half of n_1. In line with the explorative Figure 1 panels (b) and (d), for $\gamma = .5$, $\bar{k}^*$ increases, leading to similar mean estimates $\hat{\gamma}(n_1)$ and $\hat{\gamma}(\bar{k}^*)$.

We conclude that for moderate data dependences, the Zipf regression and the proposed selection procedure work well.

2 GARCH(1,1)

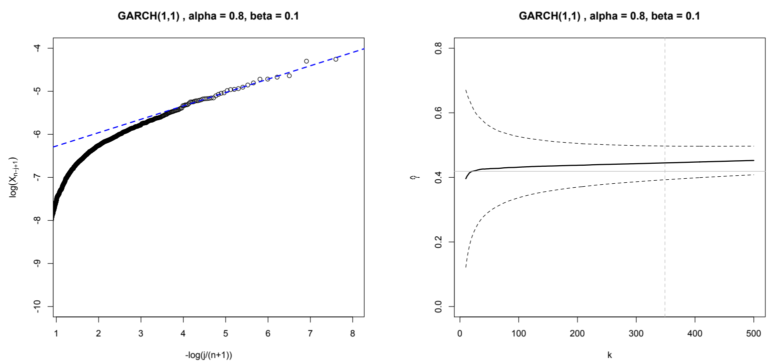
We proceed to examine the performance of the regression estimator under nonlinear dependence. The data generating process we consider is the GARCH(1,1) process

$$\begin{aligned} X_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned}$$

with $\varepsilon_t \sim N(0, 1)$. We set $\omega = 10^{-6}$ and vary the parameters $\alpha > 0$ and $\beta > 0$ such that the (weak) stationarity condition $\alpha + \beta < 1$ holds. As in Zhang et al. (2016), the population value of γ is computed by solving $E((\beta + \alpha\varepsilon_t^2)^{1/\gamma}) = 1$ numerically for γ. It can be immediately verified that the population value γ is mainly driven by α: the larger α, the larger γ. Intertemporal dependence increases with increasing α and β.

Figure 2 illustrates some results for the case $\alpha = .8$ and $\beta = .1$, based on the Monte Carlo, whilst Table 3 reports the results from the Monte Carlos for a range

Figure 2: GARCH (1,1): Pareto QQ-plot and the Zipf-regression estimates.



(a) (b)

Notes. GARCH(1,1) with $\alpha = .8$ and $\beta = .1$. Sample size $N = 10,000$, using $R = 1,000$ repetitions in the Monte Carlo study. Panel (a): One sample Pareto QQ-plot. Dashed line: Zipf-regression with slope $\hat{\gamma}(k^*)$ and threshold $X_{n-k^*,n}$ where our selection procedure determines k^*. Panel (b): Mean of estimates $\hat{\gamma}$ across the Monte Carlos (solid line), and mean 95% pointwise confidence intervals (dashed line); population value of γ (grey horizontal line), and mean k^* (dashed vertical line).

of parameterisations. The illustrative Pareto QQ-plot exhibits the by now familiar curvature. Our method selects well the ultimate linear part of the QQ-plot. Turning to the table, for lower values of α and thus of γ, the mean of k^* is about one half of n_1. As α increases, as does intertemporal dependence, k^* tends to increase as well. The mean $\hat{\gamma}(\bar{k}^*)$ is fairly close to the population value, and the mean confidence limits include the population value.

Overall, we find that the estimator performs well even for relatively strongly dependent data.

Table 3: Performance evidence: GARCH(1,1) models.

α	β	γ	n_1	mean $\hat{\gamma}(n_1)$	$\bar{k}^*$	mean $\hat{\gamma}(\bar{k}^*)$
.5	.1	.23	500	0.295	253	0.277
				(.017)	(139)	(.024)
.6	.1	.28	500	0.334	296	0.323
				(.022)	(146)	(.029)
.7	.1	.35	500	0.391	326	0.380
				(.028)	(149)	(.034)
.8	.1	.42	500	0.452	349	0.445
				(.037)	(153)	(.044)
.8	.05	.46	500	0.492	347	0.484
				(.042)	(156)	(.050)

Notes. As per Table 1. Here the DGP is a GARCH(1,1) process with stated parameters α and β.

Additional References

Gabaix, X. and R. Ibragimov (2011), "Rank - 1/2: A simple way to improve the OLS estimation of tail exponents", *Journal of Business and Economic Statistics* 29, 24-39.

Hsing, T. (1991), "On tail index estimation using dependent data", *Annals of Statistics* 19, 1547-1569.

Zhang, R., C. Li and L. Peng (2016), "Inference for the tail index of a $GARCH(1,1)$ model and an $AR(1)$ model with $ARCH(1)$ errors", *Econometric Reviews*, doi = 10.1080/07474938.2016.1224024.