

Tail-Index Estimates in Small Samples

Ronald HUISMAN

Erasmus University, 3000 DR Rotterdam, The Netherlands (r.huisman@fac.fbk.eur.nl)

Kees G. KOEDIJK

Maastricht University, 6200 MD Maastricht, The Netherlands, and CEPR (c.koedijk@berfin.unimaas.nl)

Clemens J. M. KOOL and Franz PALM

Maastricht University, 6200 MD Maastricht, The Netherlands (c.kool@algec.unimaas.nl) (f.palm@ke.unimaas.nl)

Financial returns are known to be nonnormal and tend to have fat-tailed distributions. This article presents a simple methodology that accurately estimates the degree of tail fatness, characterized by the tail index, in small samples. Our method is a weighted average of Hill estimators for different threshold values that corrects for the small-sample bias apparent in the latter. Using this estimator we produce tail-index estimates for returns on exchange rates that are close to nonbiased estimates obtained from extremely large datasets. The results indicate that many documented conclusions concerning the tail behavior of financial series are likely to have overestimated the tail fatness in small samples.

KEY WORDS: Exchange rates; Fat tails; Tail-index estimation.

It is a well-known stylized fact that financial returns tend to have empirical distributions that exhibit fatter tails than the normal distribution. [See Jansen and de Vries (1991), Koedijk, Stork, and de Vries (1992), Koedijk and Kool (1994), Loretan and Phillips (1994), and Kearns and Pagan (1997) for various recent contributions. The idea that autoregressive conditional heteroscedasticity (ARCH) effects may cause fat tails was discussed by, among others, Bera and Higgins (1993) and Bollerslev, Engle, and Nelson (1994).] The observed tail fatness of returns has important implications for economic models of asset returns. Typically, these models assume normality. Empirical measures of uncertainty with respect to financial asset movements in these models are distribution dependent. For instance, downside risk measures such as Value at Risk (VaR) focus directly on the (left) tail of the return distribution. Since the normal distribution is mostly used to fit the empirical return distribution, any discrepancy between the hypothesized normal tails and the actual fat tails potentially leads to significant errors. [Jorion (1997) and Huisman, Koedijk, and Pownall (1998) all addressed the errors in vector autoregression (VAR) estimates due to a discrepancy between assumed normality and the observed fatter tails.] If the distribution of returns is heavily nonnormal, then asset movements and speculative risks must be assessed using other measures than variance alone.

To account for tail fatness, financial returns are often modeled by a specific (nonnormal) distribution that is characterized by fat tails, like the Student- t or Stable distribution. Alternatively, generalized ARCH (GARCH) models are used to model financial returns. In the presence of (G)ARCH effects, the conditional variance of financial returns is time dependent. Periods with high conditional variances and corresponding large movements in the asset's value alternate with low-variance periods. It can be shown that such GARCH effects—even when the returns are drawn from a conditionally normal distribution—lead to an unconditional distribution with fatter tails than in the case of a time-independent unconditional normal distribution.

As argued previously, knowledge of the tail behavior of returns is of relevance in its own right. Therefore, we concen-

trate on the tail shape of the empirical distribution of financial returns. We directly estimate the tail index or maximum exponent rather than trying to fully specify the underlying true parametric distribution or to estimate the appropriate GARCH process. The tail index is a measure of the amount of tail fatness of the distribution under investigation and fits within extreme value theory (EVT). EVT investigates the distribution of the tail observations in large samples. In the limit, the tail shape follows a Pareto law for a general class of fat-tailed distributions. This limit law is characterized by the tail index, which happens to be one-to-one with the number of moments that exist. An important gain of the estimation procedure is that one can nest and test for different tail sizes. The loss consists of information about the center characteristics of the distribution. Given the predominance of outliers in asset-return series, however, one may benefit from this trade-off.

The Hill (1975) estimator is best known and most often applied, due to its easy implementation and asymptotic unbiasedness. Consequently, it has become the benchmark in the literature. However, it is biased in relatively small samples. Recently, alternative estimators have been proposed. Pictet, Dacorogna, and Müller (1996) gave an overview and studied the performance of a number of these estimators. They concluded that many estimators perform rather well for extremely large sample sizes but that all suffer severely from small-sample bias. As a result, the empirical applicability of the currently available tail-index estimators is limited to cases in which a large sample is available, either in the form of high-frequency data or in the form of a long sampling period. However, in many practical cases this condition is not fulfilled. Moreover, even when a long sample is available, it may be interesting to split the sample and analyze whether the tail structure of the sample has changed over time.

An important part of the bias in the preceding methods stems from the selection of the appropriate number of tail

observations to include in the estimation process. If one includes too many observations, the variance of the estimate is reduced at the expense of a bias in the tail estimate. This results from including too many observations from the central range. With too few observations, the bias declines but the variance of the estimate becomes too large. Application of the Hill and other estimators requires the a priori selection of the number of tail observations to include. The estimators themselves do not provide such an optimal number. [Jansen and de Vries (1991) and Koedijk and Kool (1994) used a method that requires an assumption on the underlying return distribution. Beirlant, Vynckier, and Teugels (1996, BVT) presented a method that is independent of the underlying distribution.]

In this article, we propose an alternative methodology to correct for the small-sample bias in tail-index estimates. Our method does not condition its tail estimate on one specific number of tail observations as do Hill, BVT, and other tail-index estimators. Instead, our method exploits information obtained from a set of Hill estimates each conditioned on a different number of tail observations. The result is a weighted average of a set of conventional Hill estimators, with weights obtained by simple least squares techniques.

The plan of this article is as follows. Section 1 reviews the Hill estimator and provides a theoretical motivation and explanation of the adjusted methodology. In Section 2 we present results from simulation studies to show the adequacy of our estimator with respect to different distributions and GARCH processes. It is shown that the estimator reduces the bias in Hill-based tail-index estimates dramatically for samples as small as 100 observations. Section 3 shows tail-index estimates for returns on five main U.S. dollar exchange rates. The results are compared with conventional Hill estimates. Section 4 concludes.

1. METHODOLOGY

In this section, we first briefly discuss the conventional Hill estimator, which suffers from severe small-sample bias. Subsequently, we propose a simple alternative method to obtain unbiased small-sample tail-index estimates. This estimator is based on the conventional Hill method and exploits the approximately linear relation between k and the magnitude of the bias.

1.1 The Hill Estimator

Suppose a sample of n positive independent observations is drawn from some unknown fat-tailed distribution. Let $x(i)$ be the i th-order statistic such that $x(i) \geq x(i-1)$ for $i = 2, \dots, n$. Suppose that we choose to include k observations from the right tail in our estimate. Hill (1975) proposed the following estimator for γ :

$$\gamma(k) = \frac{1}{k} \sum_{j=1}^k \ln(x(n-j+1)) - \ln(x(n-k)), \quad (1)$$

which is a maximum likelihood estimator for a conditional Pareto distribution, taking the $(k+1)$ th observation as the

threshold. The single difficulty in using Hill's estimator is the nontrivial choice of k . Extending Hall (1990), Dacarogna, Müller, Pictet, and de Vries (1995) presented an asymptotic approximation of the bias in the Hill estimator for the following class of distribution functions:

$$F(x) = 1 - ax^{-\alpha}(1 + bx^{-\beta}), \quad (2)$$

where α and β are larger than 0 and a and b are real numbers. Dacarogna et al. (1995) noted that Equation (2) provides the second-order expansion of the cumulative distribution function (cdf) for almost every fat-tailed distribution. For this class of distribution functions, Hall showed that the asymptotic expected value of the Hill estimator for a given k is approximated by

$$E(\gamma(k)) \approx \frac{1}{\alpha} - \frac{b\beta}{\alpha(\alpha+\beta)} a^{-\frac{\beta}{\alpha}} \left(\frac{k}{n}\right)^{\frac{\beta}{\alpha}}. \quad (3)$$

It is clear from Equation (3) that the bias increases in k . [This is not true for all distributions. The cdf of the Pareto distribution reads $F(x) = 1 - x^{-\alpha}$, which directly fits in (2) for a equal to 1 and b equal to 0. In this case, the bias function in (3) is 0.] However, the choice of k also affects the variance of the tail-index estimate. Hall (1990) derived the asymptotic variance of the Hill estimator for the class of distribution functions (2) as

$$\text{var}(\gamma(k)) \approx \frac{1}{k\alpha^2}. \quad (4)$$

The conclusion from Equations (3) and (4) is that a small k is preferable from the perspective of unbiasedness but a large k is preferred from an efficiency viewpoint. The trade-off between bias and precision is apparent. An important observation from the bias function (3) is that one always faces a bias for any k exceeding 0. We shall use this latter fact to present an alternative method that circumvents the selection of k .

Moreover, in line with Hall (1990), we impose the restriction $\alpha = \beta$ to approximate the asymptotic bias of the Hill estimator. Implicitly, this makes the asymptotic bias linear in k . The assumption $\alpha = \beta$ is literally true only for the limiting extreme value distribution and not for general distributions and small sample size. Dacarogna et al. (1995) stated, for instance, that in the case of Student- t distributions α equals the number of degrees of freedom of the distribution, while β equals 2. However, simulation experiments by Dacarogna et al. (1995) showed that their tail-index estimates using Hall's bootstrapping method are quite insensitive to the choice of β . They concluded, "even large errors in the assumed value of β will not lead to aberrant estimates of α ." Since actual data simulations using Student- t distributions with degrees of freedom in excess of 2 yield approximately linear bias terms over a wide range of k , imposing the constraint $\alpha = \beta$ appears warranted.

1.2 An Alternative Approach

To improve on the conventional Hill estimator, we propose to exploit an important characteristic of the bias function. For values of k smaller than a threshold value κ , the

γ estimates are seen to increase almost linearly in k justifying, indeed imposing, the restriction $\alpha = \beta$ in (3). (Unreported results show that a similar linear pattern is observed for other fat-tailed distributions like the Pareto and Burr distributions. These are available from the authors on request.) For larger k , the γ pattern depends on the exponent β/α in (3). This suggests that for k small enough the bias term can be approximated by a linear function. [We have also estimated nonlinear regression models based on the functional form for the asymptotic bias in (3). The results were much worse than those obtained by the linear specification. They confirm the appropriateness of imposing the condition $\alpha = \beta$ in (3) as was done by Hall (1990).] In this case, Equation (3) can be transformed as follows:

$$\gamma(k) = \beta_0 + \beta_1 k + \varepsilon(k), \quad k = 1, \dots, \kappa. \quad (5)$$

Instead of selecting one optimal k to estimate the tail index of the distribution under consideration, we propose to compute Hill estimates of $\gamma(k)$ for k from 1 to κ . Subsequently, the vector of computed $\gamma(k)$'s is used to estimate the parameters in Equation (5). In the previous section, we already argued that an unbiased estimate of γ could be obtained only for k approaching 0. Evaluation of Equation (5) for k approaching 0 yields an unbiased estimate of γ equal to the intercept β_0 . Applying this procedure solves the bias-variance trade-off by using the information from a whole range of conventional Hill estimates for different values of k to obtain an estimate for the tail index. κ must be chosen such that the function $\gamma(k)$ for $k = 1, \dots, \kappa$ is approximately linear. However, we will show that estimates of the tail index are quite robust with respect to the choice of κ .

1.3 Econometric Considerations

Although the parameters in (5) can be estimated using ordinary least squares (OLS), two issues complicate the procedure. First, Equation (4) indicates that the variance of Hill estimates $\gamma(k)$ is not constant for different k . The error term $\varepsilon(k)$ in Equation (5) is heteroscedastic. Therefore, we prefer a weighted least squares (WLS) approach to correct for this form of heteroscedasticity.

Second, an overlapping data problem exists due to the construction of $\gamma(k)$. The variables $\gamma(k)$ are correlated, in terms of k , since estimates $\gamma(k)$ and $\gamma(m)$, where $k \neq m$, are based on $1 + \min(k, m)$ common observations; see Equation (1). Consequently, the usual formulas for the regular standard errors both for the OLS and WLS estimates are inappropriate. Therefore, in the appendix we provide an appropriate alternative to calculate standard errors using the asymptotic normality of order statistics. Notice also that OLS and WLS do not yield fully efficient estimates of β_0 . Feasible GLS using the optimal weighting matrix (i.e., a consistent estimate of the inverse of the disturbance covariance matrix) would yield an asymptotically efficient estimate, but it has the drawback of requiring an estimate of the full error covariance matrix. (Unreported simulation results show that OLS and GLS yield estimates with similar biases for small n . In that case, the precision of GLS is lower than that of OLS. This small-sample phenomenon is

probably due to the nonlinearity of GLS. A comparison of the asymptotic standard errors of OLS and GLS for the Pareto distribution indicates that significant efficiency gains can be expected from using GLS instead of OLS in large samples.)

1.3.1 Point Estimates. For the OLS-based estimator, we first write (5) in the following matrix notation:

$$\gamma^* = Z\beta + \varepsilon, \quad (6)$$

where the vector γ^* consists of $\gamma(k)$, $k = 1, \dots, \kappa$, and Z is a $(\kappa \times 2)$ matrix with ones in the first column and the vector $\{1, 2, \dots, \kappa\}'$ in the second. As argued before, we need to correct for the heteroscedasticity in the error term ε . Equation (4) reveals that the variance of the Hill estimator is inversely related to k . We propose to apply WLS with a $(\kappa \times \kappa)$ weighting matrix W . W has $\{\sqrt{1}, \sqrt{2}, \dots, \sqrt{\kappa}\}$ as diagonal elements and zeros elsewhere. Transformation of Equation (6) through premultiplication with matrix W yields the following WLS estimate for β :

$$b_{wls} = (Z'W'WZ)^{-1}Z'W'W\gamma^* \quad (7)$$

The estimated tail index γ equals the first element of the vector b_{wls} . Due to the form of the matrices Z and W , it can be easily shown that the modified Hill estimator is a weighted average of the traditional Hill estimators for $k = 1, \dots, \kappa$:

$$\gamma^m(\kappa) = \sum_{k=1}^{\kappa} w(k)\gamma(k) \quad (8)$$

with weights $w(k)$ also depending on κ . The conventional Hill estimates $\gamma(k)$ are autocorrelated for different k due to the use of common observations. Neither OLS nor WLS estimation directly takes into account the resulting autocorrelation of the error term. Consequently, the formulas for the usual standard errors are inappropriate. In the appendix, we derive appropriate standard errors for the WLS-based estimators using the asymptotic normality of order statistics. Asymptotic properties may not hold in small samples. However, in the next section we will present evidence based on simulations that the resulting standard errors are quite accurate.

2. SIMULATION RESULTS

In this section, we investigate the statistical properties of the modified Hill estimator using simulations. In particular, we apply the modified Hill estimator to obtain tail-index estimates for relatively small samples. Observations are drawn from Student- t , Burr, and Cauchy distributions, respectively. (Results from simulations for Pareto distributions are available on request from the authors. They remain unreported because the standard Hill estimator already provides unbiased estimates for the Pareto distribution.) A similar simulation is done for a typical GARCH (1,1) process.

2.1 Student- t Distributions

In the first simulation study we draw 2,000 samples that consist of n Student- t innovations each. The sample size n equals 100, 250, 500, or 1,000 observations. The number of

degrees of freedom ranges from 1 through 5. Note that the tail index α is equal to the number of degrees of freedom of the generating Student- t distribution. For each of the 2,000 samples we calculate the tail index using the WLS-based modified Hill estimator. (OLS results differ only marginally from WLS results and therefore remain unreported. They are available on request.) In this experiment, κ equals half the sample size ($n/2$). The average over the 2,000 estimates is reported in Table 1 together with the appropriate standard errors based on the average estimate. (We have not calculated the standard errors for each single estimate. Instead, we provide the standard error based on the average estimate over the simulations. This can be done since the formulas for the standard errors only depend on the estimated tail index, the number of observations, and κ and not on any other sample characteristics that may vary over the simulation. The results of a simulation study support our use of this average estimate as an adequate estimate of the standard error.) For comparison, the standard deviations of the sampling distribution of the point estimates are reported as well.

Table 1 shows that the bias of the modified Hill estimator in small samples is small for Student- t distributions relative to the true tail index. A decrease in the sample size from 1,000 to 100 hardly affects the average value of the estimated tail index. Decreasing the sample size leads to an increase of the average standard error, however. The bias is smallest when the true α is low—that is, for relatively fat-tailed distributions. For α 's exceeding 4, the bias increases, probably due to a violation of the strong tail-fatness assumption underlying the Hill estimator. The computed standard errors are close to the standard deviation of the sampling distribution of the estimates, which supports the appropriateness of the standard-error formulas.

2.2 Cauchy and Burr Distributions

In Tables 2 and 3, we present average tail-index estimates for samples drawn from the Burr distribution and from a

Table 1. Estimates for Student- t Samples

True tail index		Sample size n			
γ	α	100	250	500	1,000
1.000	1	.940	.938	.946	.949
		(.310)	(.194)	(.136)	(.095)
.500	2	.307	.194	.140	.098
		.459	.460	.456	.456
.333	3	(.152)	(.095)	(.066)	(.046)
		.168	.106	.076	.051
.250	4	.319	.321	.316	.318
		(.105)	(.066)	(.045)	(.032)
.200	5	.126	.080	.058	.040
		.260	.258	.258	.257
		(.086)	(.053)	(.037)	(.026)
		.107	.066	.047	.034
		.227	.222	.224	.222
		(.075)	(.046)	(.032)	(.022)
		.098	.061	.031	.031

NOTE: This table provides estimates of the tails of Student- t samples for four sample sizes and five numbers of degrees of freedom (these numbers equal the true value of the tail index α of the Student- t). The numbers presented are obtained from 2,000 simulations. In each cell we present the average γ estimate over the 2,000 simulations, the standard error based on the average estimate in parentheses, and the cross-sectional standard deviation of the estimates as a benchmark for the standard error.

Table 2. Estimates for Burr Samples

True tail index		Sample size n			
γ	α	100	250	500	1,000
WLS-based modified Hill γ estimates					
.250	4	.268	.265	.265	.265
		(.088)	(.055)	(.038)	(.027)
		.100	.062	.042	.031

NOTE: This table provides γ estimates of the tails of Burr samples for four sample sizes (the true tail index α equals 4 for the Burr distribution). The numbers presented are obtained from 2,000 simulations. In each cell we present the average γ estimate over the 2,000 simulations, the corrected standard error based on the average estimate in parentheses, and the cross-sectional standard deviation of the estimates as a benchmark for the corrected standard error.

Cauchy distribution, respectively. For the Burr distribution, the cdf reads $F(x) = 1 - (x^2 + 1)^{-2}$ and α equals 4 (γ equals .25) by definition. The tail index α equals 1 for the Cauchy distribution. The performance of the modified Hill estimator is satisfactory for both distributions. The modified Hill estimator slightly underestimates when the true distribution is Cauchy. Notice that the behavior of the modified Hill estimator differs in this case from that of the traditional Hill estimator, which generally overestimates in finite samples when the true distribution is stable (e.g., see McCulloch 1997).

2.3 Choice of κ

To shed more light on the sensitivity of the previous results with respect to the choice of κ (equal to $n/2$), we present evidence for different choices of κ in Table 4. The results presented here are the average tail-index estimates over 2,000 samples with observations drawn from a Student- t distribution with degrees of freedom for seven different choices of κ smaller than $n/2$. The resulting estimates for various κ are only marginally different. We conclude that the modified Hill estimator is quite robust with respect to the choice of κ for the sample size used in this article. As a rule of thumb, we choose κ equal to $n/2$ from now on.

2.4 GARCH(1,1)

The straightforward alternative to model tail fatness in financial-return series is the use of (G)ARCH processes, in

Table 3. Estimates for Cauchy Samples

True tail index		Sample size n			
γ	α	100	250	500	1,000
WLS-based modified Hill γ estimates					
1	1	.895	.908	.911	.908
		(.295)	(.187)	(.131)	(.109)
		.258	.190	.157	.155

NOTE: This table provides γ estimates of the tails of Cauchy samples for four sample sizes (the true tail index α equals 1). The numbers presented are obtained from 2,000 simulations. In each cell we present the average γ estimate over the 2,000 simulations, the corrected standard error based on the average estimate in parentheses, and the cross-sectional standard deviation of the estimates as a benchmark for the corrected standard error.

Table 4. Estimates for Different Thresholds κ

κ	Sample size n			
	100	250	500	1,000
WLS-based modified Hill γ estimates				
$n/3$.330 (.134)	.331 (.084)	.330 (.059)	.332 (.041)
$n/4$.344 (.157)	.348 (.103)	.343 (.071)	.344 (.049)
$n^{.75}$.344 (.143)	.348 (.103)	.345 (.078)	.345 (.059)
$(\frac{n}{2})^{.9}$.342 (.138)	.344 (.091)	.342 (.066)	.344 (.048)
$(\frac{n}{2})^{.925}$.342 (.132)	.343 (.085)	.341 (.061)	.344 (.044)
$(\frac{n}{2})^{.95}$.337 (.123)	.338 (.079)	.337 (.056)	.340 (.040)
$(\frac{n}{2})^{.975}$.330 (.115)	.331 (.073)	.330 (.051)	.332 (.036)

NOTE: This table provides tail-index γ estimates obtained from the OLS-based and WLS-based modified Hill estimators. Furthermore, the threshold levels κ are varied as a function of n . The presented estimates are obtained from 2,000 simulations, in which a Student- t sample is drawn consisting of n innovations in each simulation. The true tail index α of the underlying Student- t distribution equals 3. In parentheses we present the cross-sectional standard errors.

which second-order dependence as reflected in clusters of high-volatility returns and low-volatility returns, respectively, transforms independently and identically normally distributed innovations into a fat-tailed distribution for the data. In empirical work on financial returns, the GARCH(1,1) model is commonly used; see, for instance, Baillie and Bollerslev (1989) and Drost and Nijman (1993). Consider the following standard GARCH(1,1) model:

$$y_t = \sigma_t u_t \quad (9)$$

and

$$\sigma_t^2 = \omega + \beta_0 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad (10)$$

where y is the observed return and u is a standard normally distributed innovation. Groenendijk (1999) showed that a one-to-one mapping exists between the parameters (β_0, β_1) of a GARCH(1,1) model and the tail index of the resulting distribution of returns. Since the relation is a highly nonlinear confluent hypergeometric function of the second kind, Groenendijk solved the value of the tail estimate numerically for a wide range of GARCH parameters. In Table 5, we present evidence on the performance of the (WLS-based) modified Hill estimator for a typical GARCH(1,1) model with parameters (.15, .8). These parameter values are within the range of parameter estimates that Baillie and Bollerslev (1989) reported for weekly exchange-rate returns. This particular choice is relevant and appropriate here because we will use weekly exchange-rate returns ourselves in Section 3. On the basis of the results of Groenendijk (1999), the theoretical value of the tail index α for this parameter pair is 5.78 (with corresponding γ equal to .173). For each of the sample sizes $n = 100, 250, 500$, and 1,000, we perform 2,000 replications. In the table we present the average results. Overall, the results

Table 5. Estimates for GARCH(1,1) Samples

True tail index		Sample size n			
γ	α	100	250	500	1,000
.173	5.78	.142 (.083)	.160 (.063)	.170 (.057)	.180 (.046)

NOTE: This table provides γ estimates of the tails of GARCH(1,1) samples for four sample sizes (the true tail index α equals 5.78, while the true γ equals .173) using WLS. The numbers presented are obtained from 2,000 simulations. In each cell we present the average γ estimate over the 2,000 simulations, and the cross-sectional standard deviation of the estimates as a proxy for the true standard error in parentheses. For each sample, κ is set to the half of the sample size.

are supportive of the adequate performance of the modified Hill estimator (see also Lucas 1997).

2.5 The Bias in the Conventional Hill Estimator

Finally, we compare the performance of our modified Hill estimator to the original Hill estimator in small samples. For this purpose, the simulation experiment from Table 1 is repeated for the conventional Hill estimator. This requires the choice of the number of tail observations k to include. We follow the procedure propagated by Jansen and de Vries (1991) and Koedijk and Kool (1994), among others, to determine the optimal k . [In this procedure, returns are assumed to be Student- t distributed with a specific number of degrees of freedom ($=\alpha$). Given sample size n and chosen α , a Monte Carlo experiment is performed. The mean squared error (MSE) between the estimate and the assumed α is calculated for different values of k . Then the value of k that minimizes the average MSE is chosen as the optimal number of observations to use in the Hill estimator.] BVT (1996) formulated an alternative procedure to determine k . We did not use their approach for two reasons. First, their method is quite sensitive to the precise implementation. More importantly, it is subject to the same criticism against all methods that use the conventional Hill estimator in combination with a specific procedure to determine the "optimal" k . Consequently, it is not fundamentally different from those of Jansen and de Vries (1991) and Koedijk and Kool (1994). Simulation results by Embrechts, Klüppelberg, and Mikosch (1997, pp. 197 and 337) support this view. They showed that the Hill estimator and alternatives work well over large ranges of values for k in the case of exact Pareto behavior, whereas it can lead to very wrong inference for other distributions. The "Hill horror plots" reported by Embrechts et al. (1997, p. 194) actually suggest the use of the modified Hill estimator that we propose. They show deviations of the Hill estimates trending farther away from the true value of the tail index as k is increased for other than exact Pareto distributions. The intercept from a regression through these Hill estimates is the modified Hill estimator and is close to the true value. Actually, it is much closer to the true tail-index value than the Hill estimator with k optimally chosen [using the BVT (1996) procedure or some similar procedure] over a large range of values for k .

From Table 6, we conclude that the conventional Hill estimator yields unbiased tail-index estimates for α equal to 1. For

larger α 's, the conventional Hill estimator is severely biased even if the assumption about the (unknown) underlying distribution happens to be correct. For example, when the true (and assumed) α equals 3, the average estimate of γ equals .4 for a sample size of 100; that is, the average estimate of α equals 1/.4 or 2.5. In general, therefore, the conventional Hill estimator overestimates the tail fatness of the underlying distribution. To reduce the bias of the conventional Hill estimator, it is necessary to assume a much less fat-tailed underlying distribution than the true distribution actually is. From a comparison of Tables 1 and 6, we conclude that the modified Hill estimator systematically outperforms the conventional one in small samples.

Overall, the simulation results in this section provide supporting evidence of the adequacy of the modified Hill estimator in small samples with observations drawn from Burr, Cauchy, and Student- t distributions and for returns following a GARCH(1,1) process. Moreover, the sensitivity of the modified Hill estimator for the choice of κ is low. In the next section, we will apply the modified Hill estimator to a set of weekly exchange-rate returns.

3. RESULTS ON EXCHANGE-RATE RETURNS

To illustrate the relevance of the modified Hill estimator in real-world cases, we now apply both the modified and the conventional Hill estimator to obtain tail estimates for five major exchange rates against the U.S. dollar over the period January 1979 to January 1990. The currencies considered are the French franc, the German mark, the British pound, the Swiss franc, and the Japanese yen. Using weekly returns, the sample size n equals 620. The results are in Table 7.

In the upper part of Table 7 we present the modified Hill γ estimates, while γ estimates using the conventional Hill estimator are in the lower part of Table 7. (For the procedure that we used to select the optimal k , we refer to Section 3.1. Using $n = 620$, the optimal k is found to be 20 for the analysis of one tail and 38 for the simultaneous analysis of both tails.) Again, for each exchange rate the conventional Hill procedure produces larger γ estimates than the modified Hill estimator. The conventional results imply α estimates between 3 and 4 for most exchange rates, whereas they are larger than 4 mostly if the tail index is estimated by the modified Hill estimator.

It is interesting to compare our estimates with those of Loretan and Phillips (1994), who used daily data ($n \sim 1,550$) for approximately the same sample. They applied the conventional Hill estimator to obtain tail estimates. Their use of daily data for a long period may have made the sample long enough to overcome the small-sample bias in the conventional Hill estimator. Nevertheless, for all exchange rates under consideration, our modified γ estimates suggest higher α estimates than were found by the conventional Hill estimator used by Loretan and Phillips (1994). That is, even the large number of observations they used may lead to overestimation of tail fatness using the conventional Hill estimator.

One caveat applies, though. A comparison of tail estimates at different sample frequencies is valid only as long as the observations are independent and the underlying distribution varies regularly at infinity. This is the case for many fat-tailed

Table 6. Conventional Hill Estimates

True α	Assumed α	Sample size n			
		100	250	500	1,000
$\alpha = 1$	1	1.081 (.177)	1.045 (.123)	1.038 (.087)	1.033 (.067)
	2	1.021 (.225)	1.006 (.185)	1.006 (.138)	1.008 (.104)
	3	1.010 (.296)	1.002 (.221)	1.005 (.186)	1.004 (.159)
	4	1.007 (.368)	.996 (.262)	1.003 (.222)	1.003 (.173)
	5	1.005 (.343)	.999 (.320)	1.002 (.240)	1.005 (.208)
$\alpha = 2$	1	.689 (.106)	.634 (.070)	.629 (.052)	.607 (.037)
	2	.597 (.124)	.553 (.093)	.545 (.075)	.540 (.053)
	3	.558 (.152)	.532 (.108)	.525 (.096)	.517 (.078)
	4	.538 (.181)	.522 (.129)	.519 (.110)	.514 (.084)
	5	.543 (.173)	.513 (.162)	.515 (.120)	.511 (.100)
$\alpha = 3$	1	.583 (.088)	.525 (.057)	.516 (.040)	.490 (.029)
	2	.476 (.094)	.427 (.071)	.413 (.053)	.404 (.039)
	3	.427 (.111)	.400 (.080)	.383 (.067)	.374 (.057)
	4	.400 (.134)	.386 (.094)	.371 (.078)	.370 (.062)
	5	.407 (.126)	.374 (.112)	.368 (.084)	.361 (.074)
$\alpha = 4$	1	.534 (.082)	.475 (.050)	.467 (.036)	.438 (.026)
	2	.425 (.085)	.371 (.059)	.354 (.044)	.345 (.032)
	3	.369 (.097)	.342 (.070)	.320 (.055)	.308 (.045)
	4	.341 (.112)	.326 (.079)	.307 (.063)	.302 (.049)
	5	.348 (.105)	.309 (.091)	.302 (.068)	.294 (.058)
$\alpha = 5$	1	.508 (.075)	.447 (.048)	.439 (.033)	.410 (.024)
	2	.394 (.078)	.338 (.053)	.322 (.039)	.310 (.030)
	3	.340 (.088)	.307 (.059)	.286 (.048)	.270 (.040)
	4	.312 (.100)	.288 (.067)	.273 (.055)	.264 (.044)
	5	.318 (.096)	.272 (.078)	.269 (.059)	.254 (.051)

NOTE: This table presents conventional Hill γ estimates for simulated Student- t samples. The sample size n equals 100, 250, 500, or 1,000. The true α equals either 1, 2, 3, 4, or 5. The number of tail observations (k) used in the Hill estimator is determined by a Monte Carlo simulation as described in Section 1.2. "Assumed α " refers to the degrees of freedom assumed in the Monte Carlo samples. The numbers presented are averages over 2,000 replications. The cross-sectional standard deviations of the estimates are presented in parentheses.

distributions. Then, time aggregation from daily to weekly or monthly returns does not affect the levels of the α estimates. However, with significant second-order dependence in the observations—as is the case for GARCH processes—this aggregation invariance of tail-estimate properties may be

Table 7. Tail Estimates of Foreign Exchange Rates

	France	Germany	Japan	Switzerland	U.K.
<i>Weekly returns—modified Hill γ estimates—$n \approx 620$</i>					
Both tails					
WLS	.202 (.026)	.195 (.026)	.173 (.022)	.189 (.024)	.189 (.024)
Left tail					
WLS	.166 (.031)	.168 (.032)	.121 (.024)	.137 (.026)	.202 (.038)
Right tail					
WLS	.231 (.042)	.194 (.035)	.166 (.035)	.195 (.035)	.201 (.037)
<i>Weekly returns—conventional Hill γ estimates—$n \approx 620$</i>					
Both tails					
$k = 38$.242 (.039)	.232 (.038)	.288 (.047)	.196 (.032)	.258 (.042)
Left tail					
$k = 20$.204 (.046)	.213 (.048)	.277 (.062)	.234 (.052)	.274 (.061)
Right tail					
$k = 20$.327 (.073)	.226 (.051)	.236 (.053)	.194 (.043)	.251 (.056)

NOTE: This table contains the tail-index α estimates reported by Loretan and Phillips (1994). For five exchange rates denoted against the U.S. dollar, they obtained tail-index estimates from the conventional Hill estimator (1). They obtained their estimates for five values of k (20, 30, 50, 75, 100). For each exchange rate, we report the minimum and maximum tail index they found with the standard error in parentheses and the level k at which these estimates are obtained.

violated; see Kearns and Pagan (1997), for instance. Drost and Nijman (1993) showed that the class of symmetric weak GARCH(1,1) processes is closed under temporal aggregation. That is, aggregation of weak GARCH(1,1) processes results in a weak GARCH(1,1) process again. However, the parameters of the GARCH(1,1) before and after time aggregation may differ. Consequently, the tail index may change as well, due to the result shown by Groenendijk (1999). An in-depth investigation of the relation between tail-index estimates for low- and high-frequency GARCH(1,1) processes is beyond the scope of this article and is left to future research. [In further support of the improved estimates with the modified Hill estimator in small samples, we note that Dacorogna et al. (1995), for example, applied a bootstrap methodology in combination with high-frequency data to obtain unbiased tail-index estimates. They found α estimates close to the values reported here for the same exchange rates. Again, a caveat with respect to aggregation dependence of data applies.]

4. CONCLUSION

It is a stylized fact that extreme returns on financial assets occur relatively frequently. The distribution of financial-asset returns are fatter tailed than the normal distribution, and one needs to incorporate information about the tails specifically to correctly model the shape of the distribution.

According to EVT, the shape of the distribution and tail behavior can be summarized by one characteristic parameter α , the so-called tail index. Most methods to estimate α , of which the conventional Hill (1975) estimator is best known and most widely used, suffer from small-sample bias and are

well behaved only asymptotically. Therefore, a reliable assessment of the probability of an extreme event requires a large number of observations. That is, either high-frequency data or a long sample period is required. For relatively small samples, results must be interpreted cautiously and overestimation of tail fatness is likely to be a problem.

In this article we modify the conventional Hill estimator to correct for the small-sample bias. To this end, we exploit an important characteristic of its bias function. In particular, it may be shown that the bias is an almost linear function of the number of tail observations used in the estimation. Our estimator uses a number of conventional Hill estimates, which differ in the number of tail observations included, and calculates a weighted average of these estimates with weights obtained by using simple least squares techniques. We take into account issues of heteroscedasticity and autocorrelation due to overlapping data. A procedure is developed to appropriately correct the conventional incorrect least squares standard errors.

Subsequently, we study the adequacy of our estimator in simulation studies. Overall, we conclude that approximately unbiased tail estimates result, even for samples as small as 100 observations for a range of distributions—Burr, Cauchy, and the Student- t . The modified Hill estimator appears to perform equally well for GARCH(1,1) processes, which are often used to model financial returns. We show that the modified approach is rather insensitive to the choice of maximum number of tail observations to include. Finally, the computed standard errors appear to be appropriate and similar to the standard deviations of the simulated sampling distributions. Overall, our modification provides an important gain over the conventional Hill estimator.

To illustrate the performance of the modified Hill estimator for real-world data, we apply it to obtain tail-index estimates for five main foreign exchange rates against the U.S. dollar. A comparison between the conventional and modified Hill results shows that tail fatness is estimated to be lower with the modified approach for every single exchange rate. We also try to explicitly compare our results with previously published work for the same variables and (approximately) the same period. However, these latter studies use extremely large samples (through higher-frequency data). Caution should be applied in this comparison because time aggregation may cloud the picture. The high-frequency results are generally close to our modified tail estimates for smaller (and lower-frequency samples). Still, our estimates suggest less fat tails than found even in high-frequency studies. The extent to which this is due to time aggregation issues is left to future research.

Overall, our comparative work shows that tail-fatness is easily exaggerated in small samples. Implicitly, this suggests that probabilities on extreme events may be overestimated as well when using conventional estimators. The modified Hill estimator avoids this problem and provides reliable tail-index estimates even in small samples. Our methodology can have many useful applications in real-world situations—for instance, in the risk-management industry, where samples can be relatively short while the likelihood of extreme events tends to be high. An example is the VaR- x methodology presented by Huisman et al. (1998). VaR- x directly uses tail-index estimates from

small samples to obtain Value at Risk estimates that incorporate the high probability of extreme negative returns on financial assets.

ACKNOWLEDGMENTS

We thank Michel Dacorogna, Jon Danielsson, Hans Dewachter, Piet Eichholtz, Andre Lucas, Huston McCulloch, Peter Schotman, Casper de Vries, Mark Watson, two anonymous referees, and seminar participants at Maastricht University, Erasmus University Rotterdam, and the 1997 Workshop on Finance and Econometrics in Brussels for their helpful comments. All remaining errors are our own.

APPENDIX: STANDARD ERRORS FOR THE MODIFIED HILL ESTIMATOR

In this appendix we derive the appropriate standard errors for the WLS estimator. The regular standard errors are not applicable since the nature of the modified Hill estimator presented here introduces an overlapping data problem.

Let y be the $((\kappa + 1) \times 1)$ vector of increasing order statistics $\{y(n - \kappa), \dots, y(n)\}'$ with $y(i) = \ln(x(i))$ for $i = n - \kappa, \dots, n$. Since the Hill estimator (1) is a linear combination of the $y(i)$'s, we can express the vector γ^* consisting of $\gamma(k)$ ($k = 1, \dots, \kappa$) as $\gamma^* = Ay$ for some $((\kappa \times \kappa) + 1)$ transformation matrix A . Let Σ be the covariance matrix of the order statistics contained in y . Then $\Omega = A\Sigma A'$ is the covariance matrix for the set of Hill estimates in γ^* . Once we know Ω , we obtain the standard error of the modified Hill estimates as follows. For the WLS-based estimator given in (7), the covariance matrix of b_{wls} can be obtained from

$$\text{cov}(b_{wls}) = (Z'W'WZ)^{-1}Z'W'W\Omega W'WZ(Z'W'WZ)^{-1}. \quad (\text{A.1})$$

To compute the covariance matrix in (A.1), the covariance matrix Ω must be specified or, alternatively, A and Σ must be given. The $(\kappa \times (\kappa + 1))$ matrix A can be easily derived from the conventional Hill estimator (1) for the different values of k :

$$A = \begin{bmatrix} 0 & \dots & 0 & 0 & 0 & -1 & 1 \\ 0 & \dots & 0 & 0 & -1 & 1/2 & 1/2 \\ 0 & \dots & 0 & -1 & 1/3 & 1/3 & 1/3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 1/\kappa & \dots & 1/\kappa & 1/\kappa & 1/\kappa & 1/\kappa \end{bmatrix}. \quad (\text{A.2})$$

To obtain Σ we use the fact that increasing order statistics $z(i)$ ($i = 1, \dots, \kappa + 1$) from a sample of size n are asymptotically multivariate normally distributed (e.g., see Cox and Hinkley 1974) with mean $\mu(i)$ and with covariances between order statistics $z(i)$ and $z(j)$ equal to $v(i, j)$, where

$$\mu(i) = F_z^{-1}(p(i)) \quad (\text{A.3})$$

and

$$v(i, j) = \frac{p(i)(1 - p(j))}{nf_z(\mu(i))f_z(\mu(j))} \text{ for } i \leq j. \quad (\text{A.4})$$

Here, $p(i)$ is approximated by i/n , $F_z(z)$ denotes the cdf of z , and $f_z(z)$ denotes the probability density function of z . Since any fat-tailed distribution is (approximately) Pareto distributed far in its tails, we propose to use a Pareto distribution in Equations (A.3) and (A.4) for our application. That is, we assume that the $x(i)$'s on which the modified Hill estimator is ultimately based are drawn from a Pareto distribution for $i = n - \kappa, \dots, n$. In that case, the cdf of x is given by $F(x) = 1 - x^{-\alpha}$ for $x \geq 0$. Using the fact that the Hill estimator is a linear combination of the natural logarithms of the order statistics based on x , the following expression for the approximate mean $\mu(i)$ of order statistic $y(i)$ ($i = 1, \dots, \kappa + 1$) can be found:

$$\mu(i) = \ln((1 - p(i))^{-1/\alpha}). \quad (\text{A.5})$$

Since we have an expression for $\mu(i)$ and by setting $p(i)$ in (A.4) equal to i/n and approximate α by the inverse of the estimated γ , the covariance matrix Σ of the order statistics y with elements $v(i, j)$ is fully defined. Since we have Σ , Ω is obtained from $A\Sigma A'$. Substitution in (A.3) and (A.4) leads to appropriate WLS covariance matrices.

[Received September 1997. Revised December 1999.]

REFERENCES

- Baillie, R. T., and Bollerslev, T. (1989), "The Message in Daily Exchange Rates: A Conditional-Variance Tale," *Journal of Business & Economic Statistics*, 7, 297-305.
- Beirlant, J., Vynckier, P., and Teugels, J. (1996), "Tail-index Estimation, Pareto Quantile Plots and Regression Diagnostics," *Journal of the American Statistical Association*, 91, 1659-1667.
- Bera, A. K., and Higgins, M. L. (1993), "ARCH Models: Properties, Estimation, and Testing," *Journal of Economic Surveys*, 7, 305-362.
- Bollerslev, T., Engle, and Nelson, (1994), "ARCH Models," in *Handbook of Econometrics* (vol. 4), eds. R. F. Engle and D. L. McFadden, Amsterdam: Elsevier, North-Holland, pp. 2959-3038.
- Cox, D. R., and Hinkley, D. V. (1974), *Theoretical Statistics*, London, Chapman and Hall.
- Dacorogna, M., Müller, U., Pictet, O., and de Vries, C. (1995), "The Distribution of Extremal Foreign Exchange Rate Returns in Extremely Large Data Sets," Discussion Paper 95-70, Tinbergen Institute.
- Drost, F. C., and Nijman, T. E. (1993), "Temporal Aggregation of GARCH Processes," *Econometrica*, 61, 909-927.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), *Modelling Extremal Events*, Berlin: Springer-Verlag.
- Groenendijk, P. (1999), "Essays on Exchange Rate Dynamics," unpublished PhD thesis, Free University Amsterdam, Dept. of Finance.
- Hall, P. (1990), "Using the Bootstrap to Estimate Mean Square Error and Select Smoothing Parameters in Non-parametric Problems," *Journal of Multivariate Analysis*, 32, 177-203.
- Hill, B. (1975), "A Simple General Approach to Inference About the Tail of a Distribution," *The Annals of Mathematical Statistics*, 3, 1163-1174.
- Huisman, R., Koedijk, K., and Pownall, R. (1998), "VaR-x: Fat Tails in Financial Risk Management," *Journal of Risk*, 1, 47-62.
- Jansen, D., and de Vries, C. (1991), "On the Frequency of Large Stock Returns: Putting Booms and Busts into Perspective," *The Review of Economics and Statistics*, 73, 18-24.
- Jorion, P. (1997), *Value at Risk: The New Benchmark for Controlling Market Risk*, New York: McGraw-Hill.
- Kearns, P., and Pagan, A. (1997), "Estimating the Density Tail Index for Financial Time Series," *The Review of Economics and Statistics*, 79, 171-175.

- Koedijk, K. G., and Kool, C. (1994), "Tail Estimates and the EMS Target Zone," *Review of International Economics*, 2, 153-165.
- Koedijk, K. G., Stork, P. A., and de Vries, C. G. (1992), "Differences Between Foreign Exchange Rate Regimes: The View From Tails," *Journal of International Money and Finance*, 462-473.
- Loretan, M., and Phillips, P. (1994), "Testing the Covariance Structure of Heavy-Tailed Time Series," *Journal of Empirical Finance*, 1, 211-248.
- Lucas, A. (1997), "Fat Tails in Small Samples," discussion presented at the 5th Workshop on Finance and Econometrics in Brussels, December.
- McCulloch, J. H. (1997), "Measuring Tail Thickness to Estimate the Stable Index: A Critique," *Journal of Business & Economic Statistics*, 15, 74-81.
- Pictet, O., Dacorogna, M., and Müller, U. (1996), "Hill, Bootstrap and Jackknife Estimators for Heavy Tails," working paper, Olsen & Associates.