DEUTSCHE
BUNDESBANK
EUROSYSTEM

A note on the coefficient of determination in regression models with infinite-variance variables

Jeong-Ryeol Kurz-Kim
(Deutsche Bundesbank)

Mico Loretan
(Board of Governors of the Federal Reserve System, Washington, DC)

Discussion Paper
Series 1: Economic Studies
No 10/2007

Deutsche Bundesbank, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-1
Telex within Germany 41227, telex from abroad 414431

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet http://www.bundesbank.de

ISBN 978-3–86558–294–2 (Printversion)
ISBN 978-3–86558–295–9 (Internetversion)

Abstract

Since Mandelbrot's seminal work (1963), alpha-stable distributions with infinite variance have been regarded as a more realistic distributional assumption than the normal distribution for some economic variables, especially financial data. After providing a brief survey of theoretical results on estimation and hypothesis testing in regression models with infinite-variance variables, we examine the statistical properties of the coefficient of determination in regression models with infinite-variance variables. These properties differ in several important aspects from those in the well-known finite variance case. In the infinite-variance case when the regressor and error term share the same index of stability, the coefficient of determination has a nondegenerate asymptotic distribution on the entire [0,1] interval, and the probability density function of this distribution is unbounded at 0 and 1. We provide closed-form expressions for the cumulative distribution function and probability density function of this limit random variable. In an empirical application, we revisit the Fama-MacBeth two-stage regression and show that in the infinite variance case the coefficient of determination of the second-stage regression converges to zero asymptotically.

Keywords: Regression models, alpha-stable distributions, infinite variance, coefficient of determination, Fama-MacBeth regression, Monte Carlo simulation.

JEL classifications: C12, C13, C21, G12

Nicht-technische Zusammenfassung

Die Regressionsanalyse ist eine der am häufigsten verwendeten statistisch/ökonometrischen Methoden. Dabei wird üblicherweise die Normalverteilungsannahme für die Störprozesse getroffen. Für viele empirische Daten im Finanzmarktbereich, zum Teil auch in der Makroökonomie, erweist sich die Normalverteilung aber als unzutreffend, weil deren Verteilung typischerweise wegen übermässiger Ausreißer dickere Verteilungsenden besitzt als die Normalverteilung. Als Alternative bietet sich die sogenannte alpha-stabile Verteilung an, deren Eigenschaften durch den Stabilitätsparameter (Alpha) bestimmt werden. Diese Verteilung ist besser geeignet, das Verhalten solcher Variablen zu beschreiben; die Normalverteilung ist als Spezialfall enthalten. Insofern ist die Familie der alpha-stabilen Verteilungen als eine Verallgemeinerung der (restriktiven) Normalverteilung anzusehen und nicht als Alternative.

Das Bestimmtheitsmaß ist eines der am häufigsten verwendeten Modellindikatoren. Es zeigt an, wie gut das Modell den Zusammenhang zwischen den erklärenden und erklärten Variablen beschreibt. In vielen empirischen Arbeiten wird diese Größe genutzt, um bestimmte empirische Aussagen treffen zu können. Fama und MacBeth (1973) haben beispielsweise eine sogenannte Zwei-Stufige-Regression vorgeschlagen, wobei das Bestimmtheitsmaß darüber Aufschluß gibt, wie eng Rendite und Risiko von Finanztiteln zusammenhängen. Bei Verwendung der Zwei-Stufigen-Regression haben Fama and French (1992) die empirische Aussagekraft des Capital-Asset-Pricing-Modells (CAPM) geprüft und gefolgert, dass die konventionelle Sharpe-Lintner-Version des CAPM nicht in der Lage sei, den Zusammenhang zwischen Rendite und Risiko zu erklären. Sie nahmen dabei an, dass die Renditen des US amerikanischen Aktienmarktes normalverteilt sind.

Wir untersuchen die asymptotischen Eigenschaften des Bestimmtheitsmasses im Rahmen von Regressionsmodellen. Dabei nehmen wir an, dass sowohl die Störvariable als auch die erklärende Variable alpha-stabil verteilt sind. Unsere theoretischen Ergebnisse zeigen, dass das asymptotische Verhalten des Bestimmtheitsmasses unter der alpha-stabilen Verteilungsannahme völlig anders ist als unter der Normalverteilungsannahme: die Unterschiede aus statistischer Sicht sowie in der ökonomischen Interpretation sind so groß, dass sich viele empirische Aussagen, die in der Literatur gemacht worden sind, als unzureichend, beziehungsweise falsch herausstellen. Zum Beispiel müssen die in der finanzwirtschaftlichen Literatur allgemein anerkannten empirischen Ergebnisse von Fama und French (1992) nach unserer asymptotischen Analyse grundsätzlich revidiert werden.

Non technical summary

Regression analysis is one of the most widely used methods in econometrics and statistics; the disturbances of a regression are usually assumed to be normally distributed. However, the normality assumption is not appropriate for many economic variables, especially financial market variables and also, in some cases, macroeconomic variables. Financial data are typically fat-tailed and excessively peaked around zero. Alpha-stable distributions, whose shape is governed by the stability parameter, alpha, represent one alternative distribution assumption. They are better suited to describing such financial variables; the normal distribution is a special case of the alpha-stable distribution. To that extent, the family of alpha-stable distributions is more of a generalization than an alternative.

The coefficient of determination is one of the most widely used goodness-of-fit measures for regression models. It shows how well a regression model describes the relationship between exogenous and endogenous variables. In many practical studies, this property is used to draw empirical conclusions. Fama and MacBeth (1973), for example, propose a "two-stage regression", which is designed, based on the property of the coefficient of determination, to test for the existence of a significant relationship between risk and return. Fama and French (1992, *Journal of Finance*) employ the two-stage regression to test the empirical performance of the capital asset pricing model (CAPM) and conclude that the conventional Sharpe-Lintner version of CAPM is not able to explain the relationship between risk and return. In so doing, however, they assume normality of US stock returns (CRSP data).

We study the asymptotic properties of the coefficient of determination in the framework of regression models. For our analysis, we assume that disturbances and exogenous variables alike are alpha-stable distributed. Our theoretical results show that the asymptotic behaviours of the coefficient of determination under the alpha-stable distributional assumption are far removed from those under the normality assumption: the difference is so vast – both statistically and in terms of their economic interpretation – that many empirical conclusions in the literature prove insufficient and/or invalid. For example, the mostly-cited result of Fama and French (1992) in the financial market literature, which, according to our result has no asymptotic background, should be thoroughly revised – both statistically and in terms of their economic interpretation.

Contents

A note on the coefficient of determination in regression models with infinite-variance variables[1]

1. Introduction

Granger and Orr (1972) begin their article, " 'Infinite variance' and research strategy in time series analysis," by questioning the uncritical use of the normal distribution assumption in economic modelling and estimation:

> It is standard procedure in economic modelling and estimation to assume that random variables are normally distributed. In empirical work, confidence intervals and significance tests are widely used, and these usually hinge on the presumption of a normal population. Lately, there has been a growing awareness that some economic data display distributional characteristics that are flatly inconsistent with the hypothesis of normality.

Due in part to the influential seminal work of Mandelbrot (1963), α-stable distributions are often considered to provide the basis for more realistic distributional assumptions for some economic data, especially for high-frequency financial time series such as those of exchange rate fluctuations and stock returns. Financial time series are typically fat-tailed and excessively peaked around their mean—phenomena that can be better captured by α-stable distributions with $1 < \alpha < 2$ rather than by the normal distribution, for which $\alpha = 2$.[2] The α-stable distributional assumption with $\alpha < 2$ is thus a eneralization of rather than an alternative to the Gaussian distributional assumption. If an economic series fluctuates according to an α-stable distribution with $\alpha < 2$, it is known that many of the standard methods of statistical analysis, which often rest on the asymptotic properties of sample second moments, do not apply in the conventional way. In particular, as we demonstrate in this paper, the coefficient of determination—a standard criterion for judging goodness of fit in a regression model—has several nonstandard statistical properties when $\alpha < 2$.

The linear regression model is one of the most commonly used and basic econometric tools, not only for the analysis of macroeconomic relationships but also for the study of financial market data. Typical examples for the latter case are estimation of the *ex-post* version of the capital asset pricing model (CAPM) and the two-stage modelling approach of Fama and MacBeth (1973). The first purpose of the present paper is to survey theoretical results of estimation and hypothesis testing in regression models with infinite-variance distributions, and the second is to establish that infinite variance of the regression variables has important consequences for the statistical properties of the coefficient of determination. Third, we revisit the Fama-MacBeth two-stage regression approach and demonstrate that infinite variance of the regression variables can affect decisively the interpretation of the empirical results. The rest of our paper is structured as follows. In 2 we provide a brief summary of the properties of alpha-stable distributions and of aspects of estimation, hypothesis testing, and model diagnostic checking in regression models with infinite-variance regressors and

---

[1] Jeong-Ryeol Kurz-Kim: Corresponding author. Research support from the Alexander von Humboldt Foundation is gratefully acknowledged. Deutsche Bundesbank, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main, Germany; email: jeong-ryeol.kurz-kim@bundesbank.de, Tel: +49-69-9566-4576, Fax: +49-69-9566-2982. Mico Loretan: Mailstop 18, Board of Governors of the Federal Reserve System, Washington DC 20551, USA; email: mico.loretan@frb.gov; Tel: +1-202-452-2219; Fax: +1-202-263-4850. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the staff of the Deutsche Bundesbank, the Board of Governors of the Federal Reserve System, or of any other person associated with the Federal Reserve System. We are grateful to Peter Phillips, Werner Ploberger, and Jonathan Wright for valuable comments and to Zhenyu Wang for the data used in the empirical section of this paper.

[2] The normal distribution is the only member of the family of alpha-stable distributions that has finite second (and higher-order) moments; all other members of this family have infinite variance.

disturbance terms. Section 3 provides a detailed analysis of the asymptotic distribution of the coefficient of determination in regression models with infinite-variance variables. In our empirical application, presented in Section 4, we revisit the data used in Fama and French (1992), and we show that the statistical and/or economic interpretation of their findings can be quite different under the maintained assumption of α-stable distributions from and interpretation based on the assumption of normal distributions. Section 5 summarizes the paper and offers some concluding remarks.

2 Framework

2.1 α-stable distributions

A random variable X is said to have a stable distribution if, for any positive integer $n > 2$, there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that $X_1 + \cdots + X_n \stackrel{d}{=} a_n X + b_n$, where $X_1, \ldots, X_n$ are independent copies of X and $\stackrel{d}{=}$ signifies equality in distribution. The coefficient a_n above is necessarily of the form $a_n = n^{1/\alpha}$ for some $\alpha \in (0, 2]$ (see Feller, 1971, Section VI). The parameter α is called the index of stability of the distribution, and a random variable X with index α is called α-stable. An α-stable distribution is described by four parameters and will be denoted by $S(\alpha, \beta, \gamma, \delta)$. Closed-form expressions for the probability density functions of α-stable distributions are known to exist only for three special cases.[2] However, closed-form expressions for the characteristic functions of α-stable distributions are readily available. One parameterization of the logarithm of the characteristic function of $S(\alpha, \beta, \gamma, \delta)$ is

$$\ln\left(\mathrm{E}\, e^{itX}\right) = i\delta t - \gamma^\alpha |t|^\alpha \left(1 + i\beta \operatorname{sign}(t)\, \omega(t, \alpha)\right), \tag{1}$$

where $\operatorname{sign}(t) = -1$ for $t < 0$, $\operatorname{sign}(t) = 0$ for $t = 0$, and $\operatorname{sign}(t) = +1$ for $t > 0$; and $\omega(t, \alpha) = -\tan(\pi\alpha/2)$ for $\alpha \neq 1$ and $\omega(t, \alpha) = (2/\pi) \ln|t|$ for $\alpha = 1$.

The tail shape of an α-stable distribution is determined by its index of stability $\alpha \in (0, 2]$. Skewness is governed by $\beta \in [-1, 1]$; the distribution is symmetric about δ if and only if $\beta = 0$. The scale and location parameters of α-stable distributions are denoted by $\gamma > 0$ and $\delta \in \mathbb{R}$, respectively. When $\alpha = 2$, the log characteristic function given by equation (1) reduces to $i\delta t - \gamma^2 t^2$, which is that of a Gaussian random variable with mean δ and variance $2\gamma^2$. For $\alpha < 2$ and $|\beta| < 1$, the tail behavior of an α-stable random variable X satisfies

[2] The three special cases are: (i) the Gaussian distribution $S(2, 0, \gamma, \delta) \equiv N(\delta, 2\gamma^2)$, (ii) the symmetric Cauchy distribution $S(1, 0, \gamma, \delta)$, and (iii) the Lévy distribution $S(0.5, \pm 1, \gamma, \delta)$; see Zolotarev (1986), Section 2, and Rachev et al. (2005), Section 7.

$$\lim_{x\to\infty} \mathrm{P}(X > x) = \big[C(\alpha)\,\gamma^\alpha(1+\beta)/2\big]x^{-\alpha} \quad \text{and} \tag{2}$$

$$\lim_{x\to\infty} \mathrm{P}(X < -x) = \big[C(\alpha)\,\gamma^\alpha(1-\beta)/2\big]x^{-\alpha}\,, \tag{3}$$

i.e., both tails of the probability density function (pdf) of X are asymptotically Paretian. For $\alpha < 2$ and $\beta = +1$ (-1), the distribution is maximally right-skewed (left-skewed) and only the right (left) tail is asymptotically Paretian.[3] The term $C(\alpha)$ in equations (2) and (3) is given by

$$C(\alpha) = \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)} \qquad \text{for } \alpha \neq 1 \tag{4}$$

and $2/\pi$ for $\alpha = 1$; see, e.g., Samorodnitsky and Taqqu (1994), p. 17.[4] The function $C(\alpha)$, which is shown in Figure 1, is continuous and strictly decreasing in $\alpha \in (0,2)$, with $\lim_{\alpha\downarrow 0} C(\alpha) = 1$ and $\lim_{\alpha\uparrow 2} C(\alpha) = 0$. This illustrates the well-known issue that even though all stable distributions with $\alpha < 2$ have asymptotically Paretian tails, as $\alpha \uparrow 2$ proportionately less and less of the distribution's probability mass is located in the tail region. This, in turn, requires the availability of large sample sizes in order to be able to estimate the index of stability with adequate precision when α is close to (but smaller than) 2.

Because $\mathrm{E}\,|X|^\xi = \lim_{b\to\infty}\int_0^b \mathrm{P}(|X|^\xi > x)\,\mathrm{d}x$, it follows that $\mathrm{E}\,|X|^\xi < \infty$ for $\xi \in (0,\alpha)$ and $\mathrm{E}\,|X|^\xi = \infty$ for $\xi \geq \alpha$ when X is stable with $\alpha \in (0,2)$.[5] Only moments of order up to but not including α are finite when $\alpha < 2$, and a non-Gaussian stable distribution's index of stability is also equal to its maximal moment exponent.[6] In particular, if $\alpha \in (1,2)$, the variance is infinite but the mean exists. For $\alpha > 1$, it follows that $\mathrm{E}\,X = \delta$; in addition, for $\beta = 0$, δ is equal to the distribution's mode and median irrespective of the value

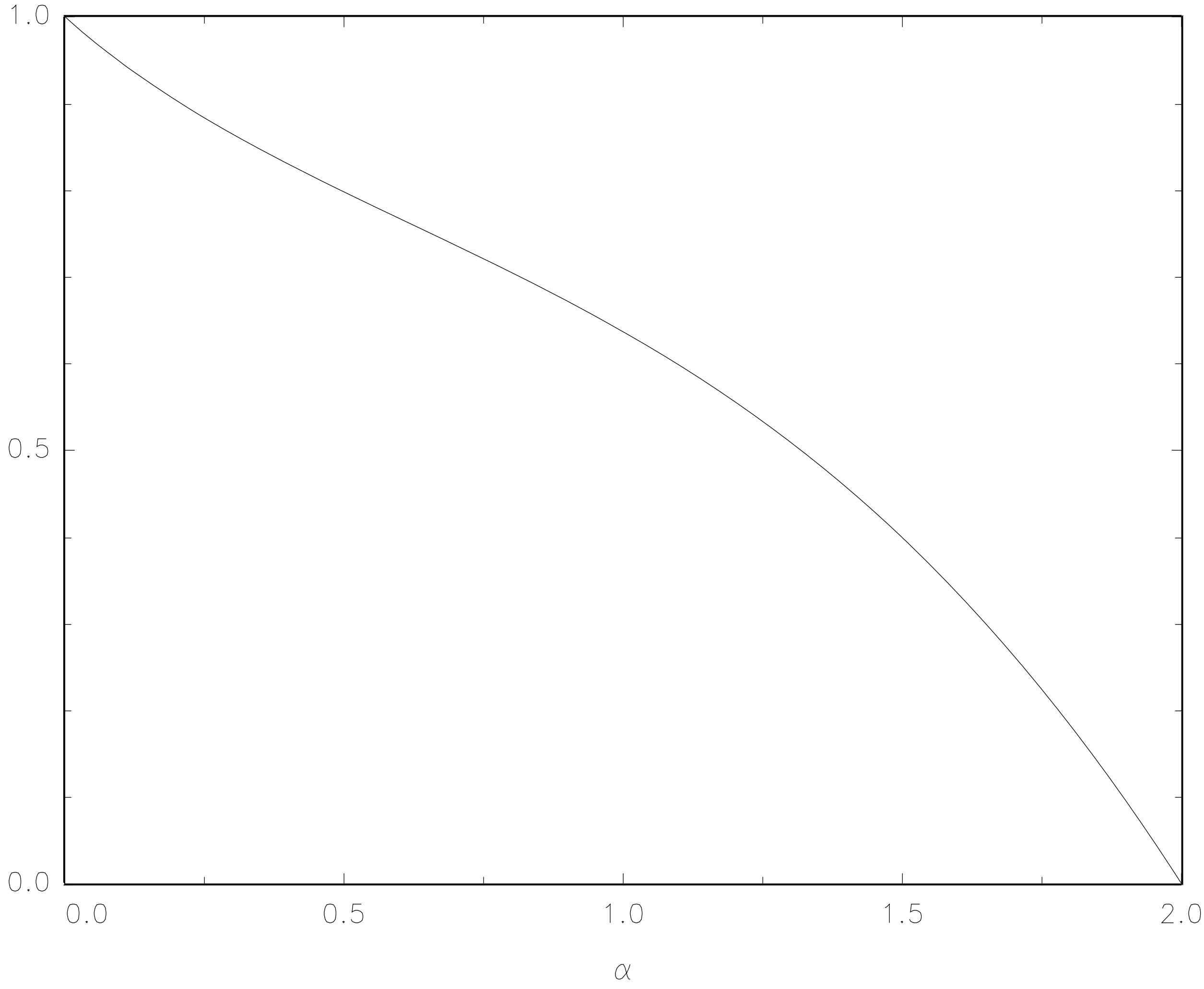
[3] For $\alpha < 1$ and $\beta = +1$, $\mathrm{P}(X < \delta) = 0$, i.e., the distribution's support is bounded below by δ. Zolotarev (1986, Theorem 2.5.3) and Samorodnitsky and Taqqu (1994, pp. 17–18) provide expressions for the rate of decline of the non-Paretian tail when $\beta = \pm 1$ and $\alpha \geq 1$.

[4] Note that the numerator and the second term in the denominator of equation (4) converge smoothly to 0 as $\alpha \to 1$; $C(1) = 2/\pi$ therefore follows from an application of L'Hôpital's Rule.

[5] Ibragimov and Linnik (1971, Theorem 2.6.4) show that this property applies not only to stable distributions, but that it pertains to *all* distributions that are in the domain of attraction of a stable distribution with index of stability α. Ibragimov and Linnik (1971, Theorem 2.6.1) provide necessary and sufficient conditions for a probability distribution to be in the domain of attraction of a stable law.

[6] The maximal moment exponent of a distribution is either a finite positive number, or it is infinite when a distribution has finite moments of all orders. For a Student-t distribution, the degrees of freedom parameter is equal to its maximal moment exponent.

Figure 1. The function $C(\alpha)$, $0 < \alpha < 2$



of α, justifying the use of the term "central location parameter" for δ in the finite-mean or symmetric cases. In addition, for $\alpha \neq 1$, one can show that $S(\alpha, \beta, \gamma, \delta) \stackrel{d}{=} \gamma \cdot S(\alpha, \beta, 1, \delta/\gamma)$.[7] We make use of this property below in the derivations of Theorem 1 and Remark 3.

The class of α-stable distributions is an interesting distributional candidate for disturbances in regression models because (i) it is able to capture the relative frequency of extreme observations in the economic variables, (ii) it has the convenient statistical property of closure under convolution, and (iii) only α-stable distributions can serve as limiting distributions of sums of independent and identically distributed (iid) random variables, as proven in Zolotarev (1986). The latter two properties are appealing for regression analysis, given that disturbances can be viewed as random variables which represent the sum of all external effects not captured by the regressors. For more details on the properties of α-stable distributions, we refer to Gnedenko and Kolmogorov (1954), Feller (1971), Zolotarev (1986), and Samorodnitsky and Taqqu (1994). The role of the α-stable distribution in financial market and econometric modelling is surveyed in McCulloch (1996) and Rachev et al. (1999).

2.2 Regression models with infinite-variance variables

Let X and Y be two jointly symmetric α-stable (henceforth, $S\alpha S$) random variables with $\alpha > 1$, i.e., we require $\mathrm{E}(X) < \infty$ and $\mathrm{E}(Y) < \infty$. Our main reason for concentrating on the case $\alpha > 1$ lies in its empirical relevance. Estimated maximal moment exponents for most empirical financial data, such as exchange rates and stock prices, are generally greater than 1.5; see, for example, de Vries (1991) and Loretan and Phillips (1994). An econometric (purposeful) reason for studying the case $\alpha > 1$ is that, for α-stable distributions with $\alpha > 1$, regression analysis that is based on sample second moments, such as least squares, is still asymptotically consistent for the regression coefficients, even though the limit distributions of these regression coefficients are nonstandard.[8] Suppose that the regression of a random variable Y on a random variable X is linear, i.e., there exists a constant θ such that

$$\mathrm{E}(\,Y \mid X\,) = \theta X \quad a.s., \tag{5}$$

with

$$\theta = \frac{[Y, X]_\alpha}{{\gamma_x}^\alpha} X,$$

[7] This result also holds for the case $\alpha = 1$ *and* $\beta = 0$, but it does not for the asymmetric Cauchy case.

[8] Another reason for this restriction comes from the viewpoint of statistical modelling. The conditional expectation of the bivariate symmetric stable distribution in (5) is, as in the Gaussian case, linear in X only when $\alpha \in (1, 2)$. The regression function is in general nonlinear, or rather only asymptotically linear, under other conditions. For more on bivariate linearity, see Samorodnitsky and Taqqu (1994, Sections 4 and 5).

where γ_x is the scale parameter of the $S\alpha S$ random variable X, and $[\cdot,\cdot]_\alpha$ is covariation (covariance in the Gaussian case), which can be calculated as $\mathrm{E}\big(XY^{<\xi-1>}\big)/\,\mathrm{E}\big(|Y|^\xi\big)$, for all $\xi \in (1,\alpha)$ with $a^{<\xi>} \equiv |a|^\xi \mathrm{sign}(a)$.

For estimation and diagnostics, the relation (5) can be written as a regression model with a constant term,

$$y_t = c + \theta x_t + u_t, \tag{6}$$

where the maintained hypothesis is that u_t is iid $S\alpha S$, with $\alpha \in (1,2]$. The econometric issues of interest are to estimate θ properly, to test the hypothesis of significance for the estimated parameter, usually based on the t-statistic, as well as to compute model diagnostics, such as the coefficient of determination, the Durbin-Watson statistic, and the F-test of parameter constancy across subsamples.

The effects of infinite variance in the disturbance term can be substantial. The ordinary least squares (OLS) estimate of θ is still consistent, but its asymptotic distribution is α-stable with the same α as the underlying variables. Furthermore, the convergence rate to the true parameter is $T^{(\alpha-1)/\alpha}$, smaller than the rate $T^{1/2}$ which applies in the finite-variance case. If $\alpha < 2$, OLS loses its best linear unbiased estimate (BLUE) property, i.e., it is no longer the minimum-dispersion estimator in the class of linear estimators of θ. In addition, the asymptotic efficiency of the OLS estimator converges to zero as the index of stability α declines to 1. Blattberg and Sargent (1971), henceforth BS, derived the BLUE for θ in (6) for the case when the value of α is known. The BS estimator is given by

$${}_\alpha\hat{\theta}_{BS} = \frac{\sum_{t=1}^T x_t^{<1/(\alpha-1)>} y_t}{\sum_{t=1}^T |x_t|^{\alpha/(\alpha-1)}}, \qquad 1 < \alpha \le 2, \tag{7}$$

which coincides with the OLS estimator when $\alpha = 2$. Kim and Rachev (1999) prove that the asymptotic distribution of the BS estimator is also α-stable. Furthermore, Kurz-Kim et al. (2006) consider an optimal power estimate based on the BS estimator for unknown α. Other efficient estimators of the regression coefficients have been studied as well; Kanter and Steiger (1974) propose an unbiased L_1-estimator, which excludes very large shocks in its estimation to avoid excess sensitivity due to outliers. Using a weighting function, McCulloch (1998) considers a maximum-likelihood estimator which is based on an approximation to a symmetric stable density. Kurz-Kim et al. (2006) provide a survey of consistent estimators of the regression coefficients for various configurations of the indices of stability of x_t and u_t.

Hypothesis testing is also affected considerably when the regressors and disturbance terms have infinite-variance stable distributions. For example, the t-statistic, commonly used to test the null hypothesis of parameter significance, no longer has a conventional Student-t distribution if $\alpha < 2$. Rather, as established by Logan et al. (1973), its pdf has modes at -1 and $+1$; for $\alpha < 1$ these modes are infinite. Kim (2003) provides empirical distributions of the t-statistic for finite degrees of freedom and various values of α by simulation. The usual applied goodness-of-fit test statistics, such as the likelihood ratio, Lagrange multiplier, and Wald statistics, also no longer have the conventional asymptotic χ^2 distribution, but have a stable χ^2 distribution, a term that was introduced by Mittnik et al. (1998).

In time series regressions with infinite-variance innovations, Phillips (1990) shows that the limit distribution of the augmented Dickey-Fuller tests for a unit root are functionals of Lévy processes, whereas they are functionals of Brownian motion processes in the finite-variance case. The F-test statistic for parameter constancy that is based on the residuals from a sample split test has an F-distribution in the conventional, finite-variance case. Kurz-Kim et al. (2005) obtain the limiting distribution of the F-test when the random variables have infinite variance. As shown by the authors, as well as by Runde (1993), the limiting distribution of the F-statistic for $\alpha < 2$ behaves completely differently from the Gaussian case: whereas in the latter case the statistic converges to 1 under the null as the degrees of freedom for both numerator and denominator approach infinity, in the former case the statistic converges to a ratio of two independent, positive, and maximally right-skewed $\alpha/2$-stable distributions. This result is used in our paper to establish the closed-form expressions for the pdf and cumulative distribution function (cdf) of the limiting distribution of the R^2 statistic when the regressor and disturbance term share the same index of stability $\alpha < 2$.

Moreover, commonly used criteria for judging the validity of some of the maintained hypotheses of a regression model, such as the Durbin-Watson statistic and the Box-Pierce Q-statistic, would be inappropriate if one were to rely on conventional critical values. Phillips and Loretan (1991) study the properties of the Durbin-Watson statistic for regression residuals with infinite variance, and Runde (1997) examines the properties of the Box-Pierce Q-statistic for random variables with infinite variance. Loretan and Phillips (1994) and Phillips and Loretan (1994) establish that both the size of tests of covariance stationarity under the null and their rate of divergence of these tests under the alternative are strongly affected by failure of standard moment conditions; indeed, standard tests of covariance stationarity are *inconsistent* when population second moments do not exist.

3 Asymptotic properties of the coefficient of determination under infinite-variance errors and infinite-variance regressors

3.1 The basic result

For the general asymptotic theory of stochastic processes with stable random variables, we refer to Resnick (1986) and Davis and Resnick (1985a, 1985b, 1986). Our results in this section are, in large part, an application of their work to the regression diagnostic context.

The maintained assumptions are:

1. The relationship between the dependent and independent variable conforms to the classical bivariate linear regression model,
$$y_t = c + \theta x_t + u_t \,. \tag{8}$$
2. u_t is iid $S\alpha S(\alpha_u, 0, \gamma_u, 0)$, with $\alpha_u \in (1, 2)$.
3. x_t is exogenous and is also iid $S\alpha S(\alpha_x, 0, \gamma_x, 0)$, with $\alpha_x \in (1, 2)$.
4. The regressor and error term have the same index of stability, i.e., $\alpha_x = \alpha_u = \alpha$.
5. The coefficients c and θ are consistently estimated by $\hat{c}$ and $\hat{\theta}$.[9]

The requirement that regressor and error term share the same index of stability is potentially rather strong. In Corollary 2 below, we examine the consequences of having different values for the indexes of stability for x_t and u_t for the asymptotic properties of the coefficient of determination.

The coefficient of determination measures the proportion of the total variation in the dependent variable that is explained by the regression. Because $\hat{y}_t - \bar{y} = \hat{\theta}(x_t - \bar{x})$ and $y_t - \bar{y} = \hat{\theta}(x_t - \bar{x}) + \hat{u}_t$, where $\bar{y}$ and $\bar{x}$ are the respective sample averages of y_t and x_t, and because $\sum_{t=1}^{T}(x_t - \bar{x})\hat{u}_t$=0 by construction, the coefficient of determination may be written as

$$\begin{aligned} R^2 &= \frac{\text{Explained Sum of Squares}}{\text{Total Sum of Squares}} \\ &= \frac{\sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \\ &= \frac{\hat{\theta}^2 \sum_{t=1}^{T}(x_t - \bar{x})^2}{\hat{\theta}^2 \sum_{t=1}^{T}(x_t - \bar{x})^2 + \sum_{t=1}^{T} \hat{u}_t^2} \,. \end{aligned}$$

Because x_t^2 and u_t^2 are in the normal domain of attraction of stable distributions with index of stability $\alpha/2$, norming by $T^{-2/\alpha}$ rather than by T^{-1} is required to obtain non-degenerate limits for the sums of the squared variables. Because $\hat{\theta} \to_p \theta$ by the assumption of consistent estimation, an application of the law of large numbers to $\bar{x}$, the continuous mapping theorem, and the results of Davis and Resnick (1985b) yield

[9] If $\alpha_u < \alpha_x$, OLS is not a consistent estimator. See Kurz-Kim et al. (2006) for a discussion of estimation methods that are consistent for various combinations of α_u and α_x.

the following expression for the joint limiting distribution of the elements in equation (8):

$$\begin{aligned}\Big(T^{-2/\alpha}\gamma_u^{-2}\sum_{t=1}^T \hat{u}_t^2\,,\,\hat{\theta}^2 T^{-2/\alpha}\gamma_x^{-2}\sum_{t=1}^T (x_t-\bar{x})^2\Big) &\sim \Big(T^{-2/\alpha}\gamma_u^{-2}\sum_{t=1}^T u_t^2\,,\,\theta^2 T^{-2/\alpha}\gamma_x^{-2}\sum_{t=1}^T x_t^2\Big)\\ &= \Big(T^{-2/\alpha}\sum_{t=1}^T (u_t/\gamma_u)^2\,,\,\theta^2 T^{-2/\alpha}\sum_{t=1}^T (x_t/\gamma_x)^2\Big)\\ &\to_d \big(S_u\,,\,\theta^2 S_x\big)\ . \end{aligned} \tag{9}$$

For $\alpha < 2$, the random variables S_u and S_x are independent, maximally right-skewed, and positive stable random variables with index of stability $\alpha/2 < 1$, $\beta = +1$, $\gamma = 1$, $\delta = 0$, and log characteristic function

$$\ln \mathrm{E}\left(e^{itS_x}\right) = \ln \mathrm{E}\left(e^{itS_u}\right) = C(\alpha)|t|^{\alpha/2}\big(1 - i\,\mathrm{sign}(t)\,\tan(\pi\alpha/4)\big)\,, \tag{10}$$

where $C(\alpha) = \big(1-\alpha/2\big)/\big(\Gamma(2-\alpha/2)\cos(\pi\alpha/4)\big)$. This leads us directly to the conclusion that the R^2 statistic of the regression model (8) has the following asymptotic limit.

Theorem 1 *Under the maintained assumptions of the model in equation* (8)*, the coefficient of determination is distributed asymptotically as*

$$R^2 \to_d \frac{\theta^2\gamma_x^2 S_x}{\theta^2\gamma_x^2 S_x + \gamma_u^2 S_u} = \frac{\eta S_x}{\eta S_x + S_u} \equiv \widetilde{R}(\alpha,\eta)\,,\ \textit{say,} \tag{11}$$

where $\eta = \theta^2\gamma_x^2/\gamma_u^2 \geq 0$.[10] *For* $\alpha < 2$*,* S_x *and* S_u *are independent and are identically distributed, have unit scale coefficient, and are strictly positive* α*-stable random variables with index of stability* $\alpha/2$*.*

Thus, for $\alpha < 2$ and $\eta > 0$, the coefficient of determination does *not* converge to a constant but has a nondegenerate distribution on the interval $[0,1]$. This contrasts strongly with the standard, finite-variance result, which is stated here for completeness.

Corollary 1 *When* x_t *and* u_t *have finite variance, the limit variables* S_x *and* S_u *in Theorem 1 are nonrandom constants and are, in fact, equal to* 1*. The limit of* R^2 *as* $T \to \infty$ *in the finite-variance case is given by*

$$R^2 \to_p \frac{\theta^2\mathrm{Var}(x)}{\theta^2\mathrm{Var}(x) + \mathrm{Var}(u)} = \frac{\eta}{\eta+1}\,,$$

where now $\eta = \theta^2\mathrm{Var}(x)/\mathrm{Var}(u)$*.*

[10] Observe that $\eta = 0$ if and only if $\theta = 0$, as the dispersion parameters γ_x and γ_u are necessarily positive.

When $\alpha < 2$, S_x and S_u are not constants, and it is this fact that causes the distribution of $\widetilde{R}$ to be nondegenerate. We postpone providing a more in-depth discussion of the intuition that underlies this result to the end of this section, i.e., after we provide a detailed analysis of the statistical properties of $\widetilde{R}(\alpha, \eta)$ when $\alpha < 2$.

Before conducting this analysis, however, we note that the fourth maintained assumption, i.e., that the regressor and error term in (8) share the same index of stability, is crucial for obtaining the result that the asymptotic distribution of $\widetilde{R}$ is nondegenerate. Indeed, when the two indices of stability differ, the following asymptotic results apply.

Corollary 2 *Suppose that the maintained assumptions of Theorem 1 apply, except that $\alpha_x \neq \alpha_u$, i.e., that the indices of stability of the regressor and error term are unequal. Suppose also that $\theta \neq 0$ to rule out the trivial case from further consideration. Then,*

- *if $\alpha_x < \alpha_u$, $1 - R^2 = o_p\left(T^{2/\alpha_u - 2/\alpha_x}\right)$; and*
- *if $\alpha_u < \alpha_x$, $R^2 = o_p\left(T^{2/\alpha_x - 2/\alpha_u}\right)$.*

Thus, R^2 converges to 1 in probability if $\alpha_x < \alpha_u$, and it converges to 0 in probability if $\alpha_u < \alpha_x$.

Proof. These results follow immediately from the fact that when $\alpha_x \neq \alpha_u$, different norming factors, *viz.*, T^{2/α_x} and T^{2/α_u}, are needed in equation (9) to achieve joint convergence of the terms $\hat{\theta} \sum_{t=1}^{T}(x_t - \bar{x})^2$ and $\sum_{t=1}^{T} \hat{u}_t^2$ to the limiting random variables S_x and S_u. When the two norming factors differ, then as $T \to \infty$ the larger factor dominates the ratio that defines R^2, and this statistic must therefore converge either to 0 or 1 in probability. E.g., suppose that $\alpha_x < \alpha_u$; then, $T^{2/\alpha_x} > T^{2/\alpha_u}$, and $T^{-2/\alpha_x} \sum \hat{u}_t^2 = T^{-2/\alpha_u}\left(T^{2/\alpha_u - 2/\alpha_x}\right) \sum \hat{u}_t^2 = o_p\left(T^{2/\alpha_u - 2/\alpha_x}\right)$. Therefore,

$$\begin{aligned} R^2 &= \frac{\hat{\theta}^2 T^{-2/\alpha_x} \sum (x_t - \bar{x})^2}{\hat{\theta}^2 T^{-2/\alpha_x} \sum (x_t - \bar{x})^2 + T^{-2/\alpha_x} \sum \hat{u}_t^2} \\ &\to_d \frac{\theta^2 \gamma_x^2 S_x}{\theta^2 \gamma_x^2 S_x + o_p\left(T^{2/\alpha_u - 2/\alpha_x}\right)} \\ &\to_p 1\,. \end{aligned}$$

Similarly, if $\alpha_u < \alpha_x$, $T^{-2/\alpha_u} \sum (x_t - \bar{x})^2 = o_p\left(T^{2/\alpha_x - 2/\alpha_u}\right)$, and R^2 converges to 0 in probability. ■ Observe that if $\alpha_u \neq \alpha_x$, the maintained assumption that the regression coefficients are estimated consistently could be relaxed, to require merely that an estimation method be employed that guarantees $\hat{\theta} \neq o_p(1)$. The result that R^2 converges either to 0 or to 1 would continue to hold in this case.

Returning to the main case of $\alpha_x = \alpha_u = \alpha$, we note that the limiting random variable $\widetilde{R}$ is defined for all values of $\alpha \in (0, 2)$. As we demonstrate later in this section, closed-form expressions for the cdf and pdf of $\widetilde{R}$ exist. We first establish some important qualitative properties of $\widetilde{R}$.

Remark 1 *For $\eta > 0$, the median m of $\widetilde{R}$ equals $\eta/(\eta + 1)$.*

Proof. For $\eta > 0$, observe that

$$\begin{aligned} \mathrm{P}\left(\widetilde{R} \leq \frac{\eta}{\eta+1}\right) &= \mathrm{P}\left(\frac{\eta S_x}{\eta S_x + S_u} \leq \frac{\eta}{\eta+1}\right) \\ &= \mathrm{P}\left(S_x \leq \frac{1}{\eta+1}\left(\eta S_x + S_u\right)\right) \\ &= \mathrm{P}\big((\eta+1)S_x - \eta S_x \leq S_u\big) \\ &= \mathrm{P}\left(S_x \leq S_u\right) . \end{aligned}$$

Because S_x and S_u are iid and have continuous cdfs, $\mathrm{P}(S_x \leq S_u) = 0.5$ by an application of Fubini's Theorem.[11] ■

Remark 2 *The distribution of $\widetilde{R}(\alpha, \eta)$ is skew-symmetric for $\eta > 0$, viz.,*

$$\widetilde{R}(\alpha, \eta) \stackrel{d}{=} 1 - \widetilde{R}(\alpha, 1/\eta) ,$$

or, equivalently, $\widetilde{R}(\alpha, m) \stackrel{d}{=} 1 - \widetilde{R}(\alpha, 1 - m)$. The pdf of $\widetilde{R}$ therefore satisfies

$$f_{\widetilde{R}(\alpha,m)}(r) = f_{\widetilde{R}(\alpha,1-m)}(1 - r) \quad \forall\ r \in [0, 1] .$$

The distribution of $\widetilde{R}$ is symmetric about 0.5 for $\eta = 1$.

Proof. Recall that S_x and S_u are iid. Thus, for $\eta > 0$

[11] See, e.g., Resnick (1999, p. 155).

$$
\begin{aligned}
1-\widetilde{R}(\alpha,1/\eta) &= 1-\frac{(1/\eta)S_x}{(1/\eta)S_x+S_u}\\
&= \frac{S_u}{(1/\eta)S_x+S_u}\\
&= \frac{\eta S_u}{\eta S_u+S_x}\\
&\overset{d}{=} \frac{\eta S_x}{\eta S_x+S_u}\\
&= \widetilde{R}(\alpha,\eta)\,.
\end{aligned}
$$

The symmetry of $\widetilde{R}$ about 0.5 for $\eta = 1$ follows immediately from this result and the fact that the distribution's support is the interval $[0, 1]$. ■

Thus, the median m of $\widetilde{R}$ is equal to the non-random limit of R^2 in the finite-variance case. However, m need not be representative of the distribution's (dominant) modes or indicate where most of the distribution's probability mass is located. As the following remark shows, the pdf of $\widetilde{R}$ has *infinite* modes at 0 and 1, i.e., at the *endpoints* of its support.

Remark 3 *For $\eta > 0$, the pdf of $\widetilde{R}$ is unbounded at 0 and 1, i.e., $f_{\widetilde{R}}(0) = f_{\widetilde{R}}(1) = \infty$. The cdf of $\widetilde{R}$ is continuous on $[0, 1]$, and the distribution does not have atoms at 0 and 1.*

Proof. To demonstrate the validity of the first part of this statement, we apply a standard result for the pdf of the ratio of two random variables,[12] adapted to the present case where the random variables in the numerator and denominator are both strictly positive. For $\eta > 0$, set $V = \eta S_x$ and $W = \eta S_x + S_u$. We have

$$
f_{\widetilde{R}}(r) = \int_0^\infty w f_{V,W}(rw, w)\,\mathrm{d}w, \qquad 0 \le r \le 1\,,
$$

where the joint pdf $f_{V,W}(\cdot,\cdot)$ is nonzero on $\mathbb{R}^+ \times \mathbb{R}^+$. The case $r = 1$ can occur only if $S_u = 0$; if $S_u = 0$, however, the random variables V and W are perfectly dependent, their joint pdf is nonzero only on the positive 45°-halfline, and the joint pdf $f_{V,V}(w, w)$ reduces to $(1/\sqrt{2})f_V(w)$, $w \ge 0$. Hence, for $r = 1$ we find

$$
f_{\widetilde{R}}(1) = \int_0^\infty w f_{V,V}(1\cdot w, w)\,\mathrm{d}w = \frac{1}{\sqrt{2}}\int_0^\infty w f_V(w)\,\mathrm{d}w = \frac{1}{\sqrt{2}}\,\mathrm{E}(\eta S_x) = \infty\,.
$$

By Remark 2, we have $f_{\widetilde{R}}(0) = \infty$ as well.

[12] See, e.g., Mood, Graybill, and Boes (1974), p. 187.

The continuity of the cdf of $\widetilde{R}$ on $[0,1]$ for $\eta > 0$ follows from the continuity of the cdfs of S_x and S_u on $\mathbb{R}^+$ and the fact that their pdfs are equal to zero at the origin. For example, one finds that $\mathrm{P}(\widetilde{R} = 1) = \mathrm{P}(S_u = 0) = 0$; the result $\mathrm{P}(\widetilde{R} = 0) = 0$ then follows from Remark 2. ■

3.2 The cdf and pdf of $\widetilde{R}$

The preceding remarks provide qualitative information about some of the unusual distributional properties of $\widetilde{R}$. However, they do not address issues such as whether the distribution has modes beyond those at 0 and 1, whether the discontinuity of the pdf at the endpoints is simple or if $f_{\widetilde{R}}(r)$ diverges (and, if so, at which rate) as $r \downarrow 0$ or $r \uparrow 1$, or how much of the distribution's mass is concentrated near the endpoints of the support. To examine these issues, we provide expressions for the cdf and pdf of $f_{\widetilde{R}}(r)$. These results are based on the following statistical properties of the ratio of two independent, positive, and maximally right-skewed stable random variables.

Proposition 1 (Zolotarev 1986, p. 205; Runde 1993, p. 11) *Let S_1 and S_2 be two positive, independent, and identically distributed α-stable random variables with common parameters $\alpha/2 \in (0,1)$, $\beta = 1$, $\gamma = 1$, and $\delta = 0$. Set $Z = S_1/S_2$. For $z \geq 0$, the cdf of Z is given by*

$$F_Z(z) = \mathrm{P}(Z \leq z) = \frac{1}{\pi\alpha/2} \arctan\left(\frac{z^{\alpha/2} + \cos(\pi\alpha/2)}{\sin(\pi\alpha/2)}\right) - \frac{1}{\alpha} + 1\,. \tag{12}$$

Differentiating this expression with respect to z, the pdf of Z for $z > 0$ is obtained as

$$f_Z(z) = \frac{\mathrm{d}}{\mathrm{d}z} F_Z(z) = \frac{\sin(\pi\alpha/2)}{\pi z\big[z^{-\alpha/2} + z^{\alpha/2} + 2\cos(\pi\alpha/2)\big]}\,. \tag{13}$$

Because Z is a positive random variable, $F_Z(z) = f_Z(z) = 0$ for $z < 0$. In addition, $\lim_{z\downarrow 0} f_Z(z) = +\infty$, and its rate of divergence is proportional to $(1/z)^{1-\alpha/2}$. Thus, the pdf of Z has a one-sided infinite singularity at 0.

Note that as $z \to \infty$, $f_Z(z) \approx \kappa \cdot z^{-\alpha/2-1}$ for a suitable constant $\kappa > 0$. This implies that Z lies in the normal domain of attraction of a stable distribution with index of stability $\alpha/2$, the same parameter as the variables S_1 and S_2. In the special case of $\alpha = 1$, S_1 and S_2 are each distributed as a Lévy stable random variable, which is well known to be equivalent to the inverse of a $\chi^2(1)$ random variable. For $\alpha = 1$, then, the pdf of Z reduces to $\big(\pi z^{1/2}(1+z)\big)^{-1}$, which is also the pdf of an $F_{1,1}$ distribution; see Runde (1993).

Because $\mathrm{P}(S_u > 0) = 1$, we may re-write $\widetilde{R}$ as

$$\begin{aligned}\widetilde{R} &= \frac{\eta S_x}{\eta S_x + S_u} \\ &= \frac{\eta (S_x/S_u)}{\eta (S_x/S_u) + 1} \\ &= \frac{\eta Z}{\eta Z + 1} = g(Z), \text{ say.}\end{aligned}$$

Note that $Z \equiv S_x/S_u$ satisfies the conditions of Proposition 1 and that the functions $g(Z)$ and $g^{-1}(\widetilde{R}) = (1/\eta)\big(\widetilde{R}/(1-\widetilde{R})\big)$ are strictly increasing and continuously differentiable in the interiors of their respective domains. We are therefore able to provide the following expressions for the cdf and pdf of $\widetilde{R}$.

Theorem 2 *For $r \in (0,1)$ and $\eta > 0$, set $z = g^{-1}(r) = (1/\eta)\big(r/(1-r)\big)$, and let the cdf and pdf of Z be given by equations (12) and (13). The cdf of $\widetilde{R}$ is given by*

$$F_{\widetilde{R}}(r) = F_Z\big[g^{-1}(r)\big], \quad 0 \le r \le 1\,. \tag{14}$$

For $r \in (0,1)$, the pdf of $\widetilde{R}$ is given by

$$\begin{aligned}f_{\widetilde{R}}(r) &= \left|\frac{\mathrm{d}}{\mathrm{d}r} g^{-1}(r)\right| \, f_Z\big[g^{-1}(r)\big] \\ &= \frac{1}{\eta(1-r)^2} \cdot \frac{\sin(\pi\alpha/2)}{\pi g^{-1}(r)\Big([g^{-1}(r)]^{-\alpha/2} + [g^{-1}(r)]^{\alpha/2} + 2\cos(\pi\alpha/2)\Big)} \\ &= \frac{\sin(\pi\alpha/2)}{\pi\, r(1-r)} \cdot \big[z^{-\alpha/2} + z^{\alpha/2} + 2\cos(\pi\alpha/2)\big]^{-1}, \quad \textit{where } z = r/\big(\eta(1-r)\big).\end{aligned} \tag{15}$$

As $r \downarrow 0$ or $r \uparrow 1$, $f_{\widetilde{R}}(r)$ diverges to ∞ at a rate proportional to $(1/r)^{1-\alpha/2}$ and $\big(1/(1-r)\big)^{1-\alpha/2}$, respectively.

Proof. The results stated in equations (14) and (15) follow immediately from Proposition 1 and the density transformation theorem.[13] Because $\lim_{r\downarrow 0} \mathrm{d}g^{-1}(r)/\,\mathrm{d}r = \eta^{-1}$, the rate of divergence of $f_{\widetilde{R}}(r)$ as $r \downarrow 0$ is equal to—apart from the multiplicative constant η^{-1}—that of $f_Z(z)$ as $z \downarrow 0$, which is $(1/z)^{1-\alpha/2}$. Finally, it follows from Remark 2 that as $r \uparrow 1$ the pdf of $\widetilde{R}$ also diverges to infinity at this rate. ■

The probability density functions and cumulative distribution functions of $\widetilde{R}(\alpha, \eta)$ for values of α between 0.25 and 1.98 are graphed in Figures 2 and 3. (In call cases, we have set $\eta = 1$.) The pdfs in Figure 2 are shown with a logarithmic scale on the ordinate. Since we know that $f_{\widetilde{R}}(0) = f_{\widetilde{R}}(1) = \infty$, we graph the functions only for $r \in \big(10^{-13}, 1 - 10^{-13}\big)$. The graphs show that

[13] See, e.g., Mood, Graybill, and Boes (1974, p. 200).

- If α is close to but less than 2, e.g., if $\alpha = 1.98$ or $\alpha = 1.90$, the pdf has an interior mode, and most of the probability mass of $\widetilde{R}$ is concentrated near its median. Conversely, only very little mass is located near 0 and 1, and the pdfs register only mild increases as r approaches either edge of the distribution's support.

- For $\alpha = 1.75$ and $\alpha = 1.50$, the distribution of $\widetilde{R}$ continues to have an interior mode (as well as, of course, the two unbounded modes at 0 and 1). However, the distribution is noticeably less concentrated around the interior mode than when α is closer to 2.

- By $\alpha = 1.20$, the interior mode has disappeared and the distribution is nearly uniform over the entire interval $[0, 1]$.

- When α takes on even smaller values, less and less probability mass of $\widetilde{R}$ is located near the median, and more and more of the probability mass becomes concentrated at 0 and 1. E.g., when $\alpha = 0.25$, about 75 percent of the probability mass lies within 0.001 of the two endpoints of the distribution, while the probability of observing a realization of $\widetilde{R}$ for $r \in [0.25, 0.75]$ is less than 5 percent.

A heuristic explanation of these properties of R is straightforward. We begin by recalling that the term $C(\alpha)$, shown in equation (4) and Figure 1, factors into the probability of obtaining very large values of the random variables in question; if α is close to 2, $C(\alpha)$ is close to 0, and the fraction of observations of x_t and u_t that fall into the Paretian-tail region is therefore very low. This low probability of observing large observations of x_t and u_t, in turn, makes it unlikely to observe a very large draw of either S_x or S_u, i.e., of generating an observation of R that is either close to 1 or close to 0. As a result, if α is close to 2, most of the mass of R is concentrated near its median. As α moves down and away from 2, say to around 1.5, the probability of observing tail observations for x_t and u_t increases, leading to a higher frequency of observing very large values of S_x and S_u as well; in consequence, the central mode of R around m becomes less pronounced and more of its probability mass is located near the edges of its support. Finally, as α decreases further, the central mode vanishes entirely because it is more and more likely that some extremely large observations of x_t or u_t will influence the realization of R. In the limit, as $\alpha \downarrow 0$, R degenerates to a Bernoulli random variable, with all of its mass concentrated at 0 and 1.

Figure 2. Probability density functions of $\widetilde{R}(\alpha,\eta)$

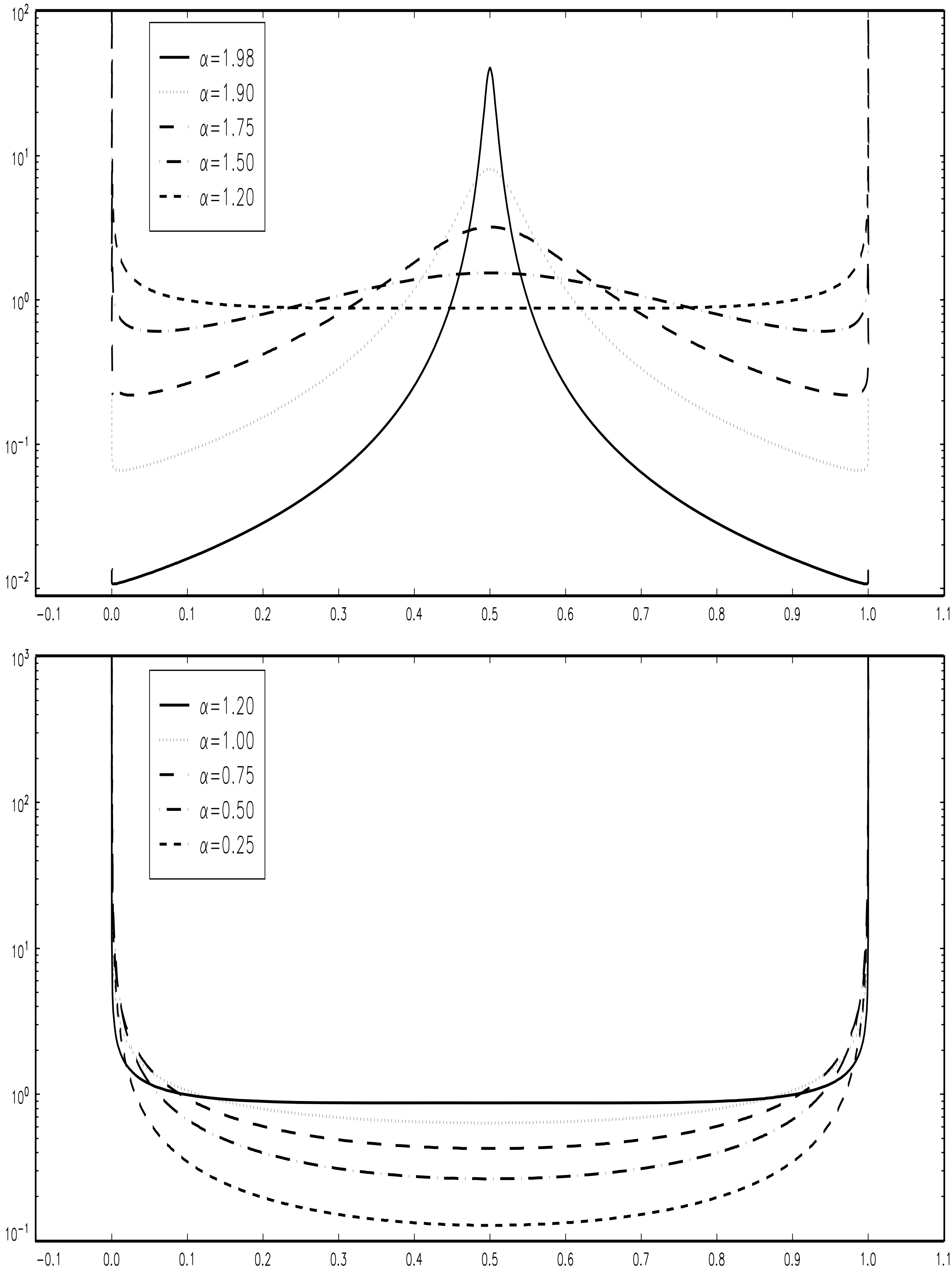
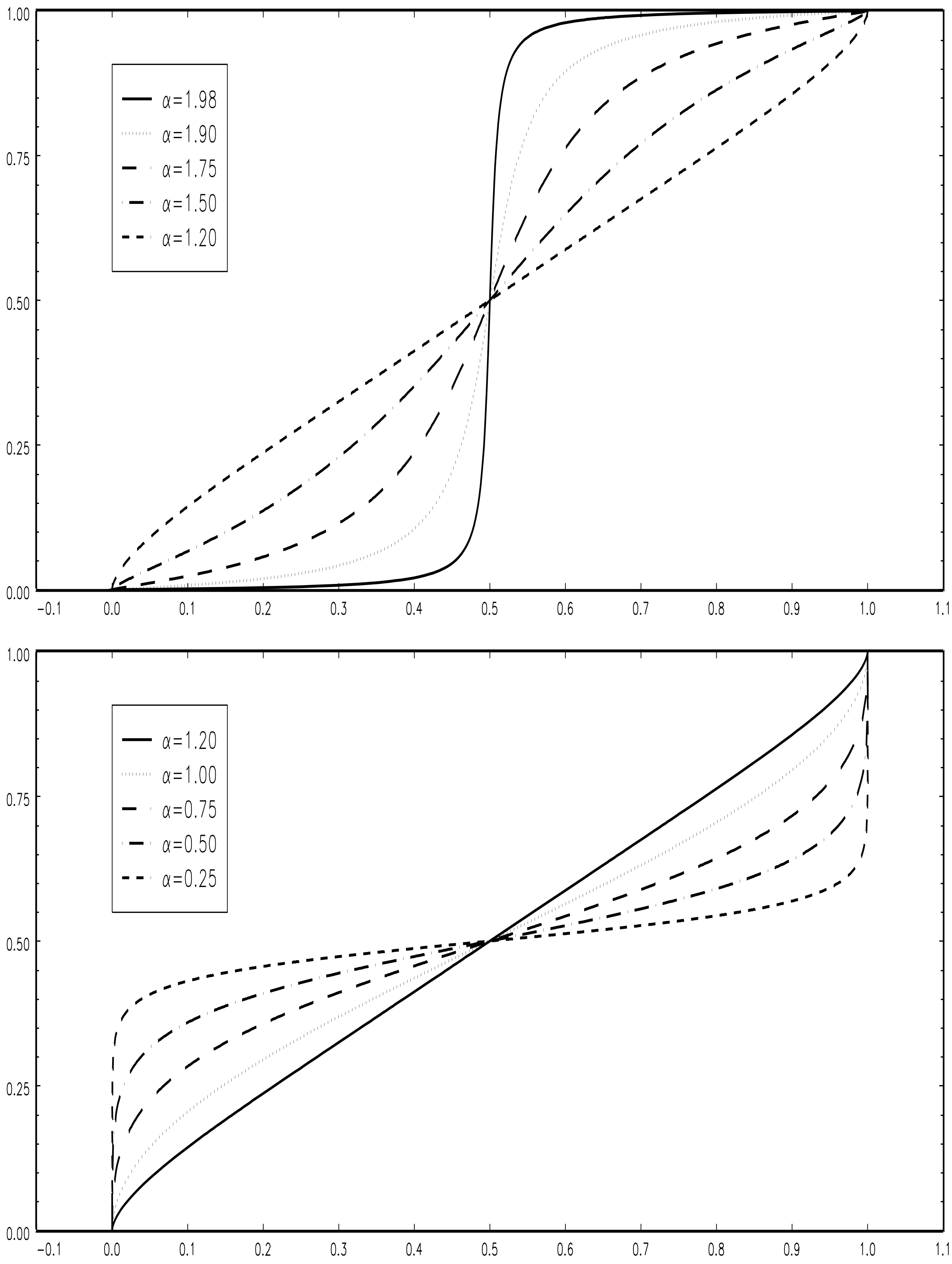


Figure 3. Cumulative distribution functions of $\widetilde{R}(\alpha, \eta)$



4 An empirical application

Fama and MacBeth (1973) proposed the so-called Fama-MacBeth regression to test the hypothesis of a linear relationship between risk and risk premium in stock returns in a cross-sectional setting. Let r_{it} be the return on the i-th market portfolio at time t. The whole return of the market portfolio with cross-sectional dimension N and time dimension T can be written in matrix form as

$$r := \left[r_{it}\right]_{n=1,\ldots,N;\, t=1,\ldots,T} = \begin{bmatrix} r_{1,1} & \cdots & r_{1,T} \\ \vdots & \ddots & \vdots \\ r_{N,1} & \cdots & r_{N,T} \end{bmatrix}.$$

Define $\bar{r}_i := T^{-1}\sum_{t=1}^{T} r_{it}$, $i = 1,\ldots,N$; $R_t := N^{-1}\sum_{i=1}^{N} r_{it}$, $t = 1,\ldots,T$; and $\mu_R = T^{-1}\sum_{t=1}^{T} R_t$. The first-stage Fama-MacBeth regression is an *ex post* CAPM,

$$r_{it} = \theta_{0i} + \theta_i R_t + u_t, \quad t = 1,\ldots,T, \tag{16}$$

where $\mathrm{E}\, u_t = 0$, $\mathrm{E}(u_t R_t) = 0$, and u_t is iid $S\alpha S$ with the same $\alpha \in (1,2]$ as r_{it}. We may assume that the distribution of θ_i has a finite mean and variance. Denote the OLS estimates of the regression coefficients in equation (16) by $\hat{\theta}_i$ and $\hat{\theta}_{0i}$. The second-stage Fama-MacBeth regression is given by

$$\bar{r}_i = \gamma_0 + \gamma_1 \hat{\theta}_i + e_i, \qquad i = 1,\ldots,N\,, \tag{17}$$

where e_i is iid $S\alpha S$ with the same α as r_{it}, $\mathrm{E}\, e_i = 0$, and $\mathrm{E}(e_i\hat{\theta}_i) = 0$. Again, γ_1 and γ_0 in the second-stage Fama-MacBeth regression can be estimated by OLS. The R^2 statistic of the second-stage Fama-MacBeth regression is given by

$$R^2 = \frac{N^{-1}\hat{\gamma}_1^2\sum_{i=1}^{N}\big(\hat{\theta}_i - \bar{\hat{\theta}}\big)^2}{N^{-1}\hat{\gamma}_1^2\sum_{i=1}^{N}\big(\hat{\theta}_i - \bar{\hat{\theta}}\big)^2 + N^{-1}\sum_{i=1}^{N}\hat{e}_i^2}\,. \tag{18}$$

The following theorem states the limiting values of R^2 in the finite-variance and the infinite-variance cases.

Theorem 3 *If the returns r_{it} follow an iid α-stable distribution with $\alpha \in (1,2]$ and if $\mu_R > 0$, the R^2 statistic in* (18) *has the following limits as $T \to \infty$ and $N \to \infty$:*

- *If $\alpha = 2$, $R^2 \to_p \gamma_1^2\mathrm{Var}(\theta_i)/\big(\gamma_1^2\mathrm{Var}(\theta_i) + \mathrm{Var}(\varepsilon)\big)$; and*
- *If $\alpha < 2$, $R^2 = o_p\big(N^{1-2/\alpha}\big)$.*

Thus, when $\alpha < 2$, $R^2 \to_p 0$, at a rate that is proportional to $N^{1-2/\alpha}$.

Proof. The result for the finite-variance case follows immediately from Corollary 1. For $\alpha < 2$, observe that the normalized estimator of θ_i, $T^{(\alpha-1)/\alpha}\left(\hat{\theta}_i - \theta_i\right)$, is in the domain of attraction of an α-stable distribution for finite values of T. As $T \to \infty$, the dispersion of $\hat{\theta}_i$ about θ_i therefore converges to 0, and the distributional properties of the regressor $\hat{\theta}_i$ in the second-stage Fama-MacBeth regression converge to those of θ_i; by assumption, the variance of θ_i is finite. Thus, as $N \to \infty$ and $T \to \infty$, the numerator in equation (18) converges to $\gamma_1^2 \mathrm{Var}(\theta_i)$. In contrast, the second summand in the denominator of (18) requires norming by $N^{2/\alpha} > N$ in order to attain a proper limit. The coefficient of determination therefore converges to 0 as $N \to \infty$ and $T \to \infty$, at a rate equal to $N^{1-2/\alpha}$. ■

This result does not conflict with the one provided in Theorem 1, as the present case is one of an unbalanced regression design: the regressor has an asymptotically finite variance, whereas the error term has infinite variance. Instead, this result is closely related to the one provided in Theorem 2, which examined the asymptotic limit of R^2 when $\alpha_x \neq \alpha_u$. We note that even if T is fixed (as is generally taken to be the case in Fama-MacBeth regressions), the dispersion of $\hat{\theta}_i$ will likely be quite a bit smaller than that of e_i, indicating that the expected value of η, and hence the expected value of the median of R^2, in the second-stage regression will be quite small unless γ_1^2 is sufficiently large. These qualitative observations are confirmed by a small Monte Carlo simulation, shown in Table 1, in which we report the median value of R^2 as a function of α, N, and T. In particular, it is evident for $\alpha = 1.5$ and $\alpha = 1.75$ that the median value of R^2 declines as N increases when T is fixed, and that this effect is particularly pronounced when T is large.

On the basis of the small value of coefficient of determination from the Fama-MacBeth regression, Jagannathan and Wang (1996) confirm the finding of Fama and Macbeth (1973) of a "flat" relation between average return and market beta. They report a very low coefficient of determination of 1.35%=0.0135 for the Sharpe-Lintner-Black (SLB) static CAPM. Regarding "thick-tailed" phenomena in empirical data, Fama and French (1992) conjectured that neglecting the heavy-tails phenomenon of the data does not lead to serious errors in the interpretation of empirical results.

In the following, we use the same CRSP dataset as was used by Jagannathan and Wang (1996); the data are very similar to those that were used by Fama and French (1992). The data consist of stock returns of nonfinancial firms listed on the NYSE and AMEX from 1962 until 1990 covered by CRSP alone. For our analysis we generalize the implicit assumption of normality to that of α-stable distributions. The point

Table 1: Coefficient of determination R^2 as a function of T, N, α, and μ_R

T	N	α	μ_R				
			0.0	0.1	0.3	0.5	1.0
100	30	1.5	0.0404	0.0425	0.0576	0.1009	0.2963
		1.75	0.0488	0.0543	0.1332	0.2944	0.6410
		2	0.0613	0.0820	0.3064	0.5525	0.8316
	100	1.5	0.0162	0.0172	0.0292	0.0596	0.2019
		1.75	0.0260	0.0328	0.1032	0.2413	0.5756
		2	0.0420	0.0640	0.2829	0.5223	0.8140
	500	1.5	0.0068	0.0075	0.0147	0.0325	0.1206
		1.75	0.0177	0.0222	0.0779	0.1941	0.5055
		2	0.0374	0.0587	0.2749	0.5131	0.8083
	1000	1.5	0.0046	0.0055	0.0114	0.0264	0.0973
		1.75	0.0149	0.0199	0.0720	0.1778	0.4792
		2	0.0371	0.0610	0.2768	0.5140	0.8082
250	30	1.5	0.0402	0.0426	0.0779	0.1598	0.4417
		1.75	0.0474	0.0642	0.2509	0.4899	0.7993
		2	0.0608	0.1237	0.5253	0.7542	0.9248
	100	1.5	0.0161	0.0190	0.0448	0.1058	0.3304
		1.75	0.0265	0.0430	0.2066	0.4264	0.7560
		2	0.0428	0.1049	0.4965	0.7328	0.9164
	500	1.5	0.0064	0.0075	0.0220	0.0565	0.2020
		1.75	0.0169	0.0290	0.1571	0.3500	0.6950
		2	0.0385	0.1027	0.4893	0.7260	0.9137
	1000	1.5	0.0047	0.0058	0.0172	0.0452	0.1667
		1.75	0.0143	0.0251	0.1440	0.3273	0.6730
		2	0.0352	0.1002	0.4860	0.7239	0.9128
1000	30	1.5	0.0387	0.0484	0.1499	0.3320	0.6748
		1.75	0.0470	0.1193	0.5351	0.7665	0.9309
		2	0.0592	0.3340	0.8175	0.9254	0.9802
	100	1.5	0.0162	0.0223	0.0940	0.2272	0.5558
		1.75	0.0265	0.0871	0.4612	0.7124	0.9115
		2	0.0434	0.3041	0.7988	0.9167	0.9778
	500	1.5	0.0065	0.0104	0.0521	0.1341	0.3994
		1.75	0.0169	0.0635	0.3910	0.6507	0.8865
		2	0.0387	0.2995	0.7920	0.9136	0.9768
	1000	1.5	0.0046	0.0072	0.0399	0.1079	0.3443
		1.75	0.0144	0.0579	0.3663	0.6257	0.8744
		2	0.0362	0.2964	0.7908	0.9131	0.9767
2500	30	1.5	0.0403	0.0580	0.2478	0.4806	0.7962
		1.75	0.0480	0.2185	0.7214	0.8804	0.9677
		2	0.0589	0.5557	0.9179	0.9688	0.9920
	100	1.5	0.0155	0.0294	0.1621	0.3581	0.6973
		1.75	0.0255	0.1704	0.6599	0.8474	0.9578
		2	0.0418	0.5256	0.9086	0.9650	0.9910
	500	1.5	0.0066	0.0130	0.0883	0.2243	0.5507
		1.75	0.0169	0.1251	0.5900	0.8066	0.9452
		2	0.0372	0.5136	0.9047	0.9634	0.9906
	1000	1.5	0.0047	0.0103	0.0737	0.1896	0.4970
		1.75	0.0149	0.1202	0.5674	0.7902	0.9394
		2	0.0370	0.5163	0.9045	0.9633	0.9906

The numbers in the table are the median R^2 from each simulated distribution with 100,000 replications.

estimate of α for stock returns the CRSP data, using both the Hill method (Hill, 1975) and the Dufour and Kurz-Kim Monte Carlo method (Dufour and Kurz-Kim, 2006), is $\hat{\alpha} = 1.86$. McCulloch's quantile estimation method (McCulloch, 1986) produces $\hat{\alpha} = 1.89$. The total return for the sample period is $\hat{\mu}_R = 0.1088$. The simulated median R^2, from 100,000 replications, is 0.3823 for $\alpha = 2$, but it is only 0.0940 for $\alpha = 1.86$. The empirical cumulative probability of obtaining a coefficient of determination of 0.0135 is a minuscule 1.28% for $\alpha = 2$, but it is a more sizable 18.61% for $\alpha = 1.86$. Thus, the inference drawn from the low value of R^2 by Fama and French (1992)—that the empirical usefulness of the SLB CAPM is refuted—does not seem to be robust once proper allowance is made for the distributional properties of the data that give rise to this statistic.

5 Concluding remarks

After providing a brief overview of some of the properties of α-stable distributions, this paper surveys the literature on the estimation of linear regression models with infinite-variance variables and associated methods of conducting hypothesis and specification tests. Our paper adds to the already-wide body of knowledge that there are substantial differences between regression models with infinite-variance and finite-variance regressors and error terms by examining the properties of the coefficient of determination. In the infinite-variance case with iid regressors and error terms that share the same index of stability α, we find that the R^2 statistic does not converge to a constant but instead that it has a nondegenerate asymptotic distribution on the $[0, 1]$ interval, with a pdf that has infinite singularities at 0 and 1. We provide closed-form expressions for the cdf and pdf of this limit random variable. Finally, we provide an application of our methods to the Fama-MacBeth two-stage regression setup and show that the coefficient of determination asymptotically converges to 0 in probability when the regression variables have infinite variance.

Given the random nature and some of the unusual properties of the limit law $\widetilde{R}$ when the regressors and error terms share the same index of stability, and given our related finding that R^2 converges to zero in probability when the tail index of the disturbance term is smaller than that of the regressor, we view our results as establishing that one should *not* rely on R^2 as a measure of the goodness of fit of a regression model whenever the regressors and disturbance terms are sufficiently heavy-tailed to call into question the existence of second (population) moments. At the very least, if one insists on reporting the coefficient of determination in regressions with infinite-variance variables, one should also report a point estimate of the median of R^2, $\hat{m} = \hat{\eta}/(\hat{\eta} + 1)$, where η is as in Theorem 1, and indicate whether the error terms and regressors may reasonably be assumed to share the same index of stability. It is widely known—and is certainly stressed in all introductory econometrics textbooks—that a *high* value of R^2 does not provide a sufficient basis for concluding that an empirical regression model is a "good" explanation of the dependent variable, or even that the regression is correctly specified. Nevertheless, one suspects, researchers may view empirical regressions with *low* values of R^2 as an indication that the (linear) regression relationship is either weak or unreliable. A direct implication of the work presented in this paper is that whenever the data are characterized by significant outlier activity, a low value of R^2 should not by itself be used to disqualify the model from further consideration.

Several extensions to the work presented here are possible. First, it seems desirable to study how well the distribution of $\widetilde{R}$ approximates the empirical distribution of R^2 in finite samples, for various types of heavy-tailed distributions and for various types of estimators (including OLS, Blattberg-Sargent's BLUE, and the least-absolute deviation estimator). Second, the theoretical results presented in our paper depend crucially on the assumption that the random variables are iid. Relaxing this assumption would seem to be useful, as many economic time series have interesting dependence and heterogeneity features. Introducing serial dependence and heterogeneity, especially conditional heterogeneity, in our model would serve the purpose of studying how the properties of $\widetilde{R}$ may be affected by such departures from the basic case of iid variables. The authors are considering conducting research to extend the work presented in this paper along these lines.

References

Blattberg, R. and T. Sargent (1971): Regression with non-Gaussian stable disturbances: Some sampling results, *Econometrica* 39, 501–510.

Davis, R.A. and S.I. Resnick (1985a): Limit theory for moving averages of random variables with regularly varying tail probabilities, *Annals of Probability* 13, 179–195.

Davis, R.A. and S.I. Resnick (1985b): More limit theory for sample correlation functions of moving averages, *Stochastic Processes and Their Applications* 20, 257–279.

Davis, R.A. and S.I. Resnick (1986): Limit theory for sample covariance and correlation functions of moving averages, *Annals of Statistics* 14, 533–558.

De Vries, C. (1991): On the relation between GARCH and stable processes, *Journal of Econometrics* 48, 313–324.

Dufour, J.-M. and J.-R. Kurz-Kim (2006): Exact inference and optimal invariant estimation for the stability parameter of symmetric α-stable distributions, *Journal of Empirical Finance*, forthcoming.

Fama, E.F. and K.R. French (1992): The cross-section of expected stock returns, *The Journal of Finance* 47, 427–465.

Fama, E.F. and J.D. MacBeth (1973): Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 71, 607–636.

Feller, W. (1971): *An Introduction to Probability Theory and its Applications*, Vol. 2 (2nd ed.). Wiley: New York.

Gnedenko, B.V. and A.N. Kolmogorov (1954): *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley: Reading, Mass.

Granger, C.W.J. and D. Orr (1972): "Infinite variance" and research strategy in time series analysis, *Journal of the American Statistical Association* 67, 275–285.

Ibragimov, I.A. and Yu.V. Linnik (1971): *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff: Groningen.

Hill, B.M. (1975): A simple general approach to inference about the tail of a distribution, *Annals of Statistics* 3, 1163–1174.

Jagannathan, R. und Z. Wang (1996): The conditional CAPM and cross-section of expected returns, *The Journal of Finance* 51, 3–53.

Kanter, M. and W.L. Steiger (1974): Regression and autoregression with infinite variance, *Advances in Applied Probability* 6, 768–783.

Kim, J.-R. (2003): Finite-sample distributions of self-normalized sums, *Computational Statistics* 18, 493–504.

Kim, J.-R. and S.T. Rachev (1999): Asymptotic distributions of BLUE in the presence of infinite variance residuals, Unpublished manuscript.

Kurz-Kim, J.-R., S.T. Rachev and G. Samorodnitsky (2005): F-type Distributions for Heavy-tailed Variates, Unpublished manuscript.

Kurz-Kim, J.-R., S.T. Rachev, G. Samorodnitsky and S. Stoyanov (2006): Asymptotic distribution of linear unbiased estimators in presence of heavy-tailed stochastic regressors and residuals, *Probability and Mathematical Statistics*, forthcoming.

Logan, B.F., C.L. Mallows, S.O. Rice and L.A. Shepp (1973): Limit distributions of self-normalized sums, *Annals of Probability* 1, 788–809.

Loretan, M. and P.C.B. Phillips (1994): Testing the covariance stationarity of heavy-tailed time series, *Journal of Empirical Finance* 1, 211–248.

Mood, A.M., F.A. Graybill and D.C. Boes (1974): *Introduction to the Theory of Statistics*, 3rd ed. McGraw-Hill: New York.

Mandelbrot, B. (1963): The variation of certain speculative prices, *Journal of Business* 36, 394–419.

McCulloch, J.H. (1986): Simple consistent estimators of stable distribution parameters, *Communications in Statistics, Computation and Simulation* 15, 1109–1136.

McCulloch, J.H. (1996): Financial applications of stable distributions, in: *Handbook of Statistics—Statistical Methods in Finance* Vol. 14, 393–425, eds. G.S. Maddala and C.R. Rao. Elsevier Science B.V.: Amsterdam.

McCulloch, J.H. (1998): Linear regression with stable disturbances, in: *A Practical Guide to Heavy Tails*, eds. R. Adler, R. Feldman, and M.S. Taqqu. Birkhäuser: Berlin and Boston.

Mittnik, S., S.T. Rachev and J.-R. Kim (1998): Chi-square-type distributions for heavy-tailed variates, *Econometric Theory* 14, 339–354.

Phillips, P.C.B. (1990): Time series regression with a unit root and infinite-variance errors, *Econometric Theory* 6, 44–62.

Phillips, P.C.B. and M. Loretan (1991): The Durbin-Watson ratio under infinite-variance errors, *Journal of Econometrics* 47, 85–114.

Phillips, P.C.B. and M. Loretan (1994): On the theory of testing covariance stationarity under moment condition failure, in: *Advances in Econometrics and Quantiative Economics. Essays in Honor of Professor C.R. Rao*, 198–233, eds. G.S. Maddala, P.C.B. Phillips, and T.N. Srinivasan. Blackwell: Oxford.

Rachev, S.T., J.-R. Kim and S. Mittnik (1999): Stable Paretian econometrics, Part I and II, *The Mathematical Scientists* 24, 24–55 and 113–127.

Rachev, S.T., C. Menn and F.J. Fabozzi (2005): *Fat-Tailed and Skewed Asset Return Distributions*. Wiley: New York.

Resnick, S.I. (1986): Point processes regular variation and weak convergence, *Advances in Applied Probability* 18, 66–138.

Resnick, S.I. (1999): *A Probability Path*. Birkhäuser: Berlin and Boston.

Runde, R. (1993): A note on the asymptotic distribution of the F-statistic for random variables with infinite variance, *Statistics & Probability Letters* 18, 9–12.

Runde, R. (1997): The asymptotic null distribution of the Box-Pierce Q-statistic for random variables with infinite variance, with an application to German stock returns, *Journal of Econometrics* 78, 205–216.

Samorodnitsky, G. and M.S. Taqqu (1994): *Stable Non-Gaussian Random Processes*. Chapman & Hall: New York.

Zolotarev, V.M. (1986): *One-Dimensional Stable Distributions*, Translations of Mathematical Monographs, Vol. 65. American Mathematical Society: Providence.

The following Discussion Papers have been published since 2006:

Series 1: Economic Studies

1	2006	The dynamic relationship between the Euro overnight rate, the ECB's policy rate and the term spread	Dieter Nautz Christian J. Offermanns
2	2006	Sticky prices in the euro area: a summary of new micro evidence	Álvarez, Dhyne, Hoeberichts Kwapil, Le Bihan, Lünnemann Martins, Sabbatini, Stahl Vermeulen, Vilmunen
3	2006	Going multinational: What are the effects on home market performance?	Robert Jäckle
4	2006	Exports versus FDI in German manufacturing: firm performance and participation in international markets	Jens Matthias Arnold Katrin Hussinger
5	2006	A disaggregated framework for the analysis of structural developments in public finances	Kremer, Braz, Brosens Langenus, Momigliano Spolander
6	2006	Bond pricing when the short term interest rate follows a threshold process	Wolfgang Lemke Theofanis Archontakis
7	2006	Has the impact of key determinants of German exports changed? Results from estimations of Germany's intra euro-area and extra euro-area exports	Kerstin Stahn
8	2006	The coordination channel of foreign exchange intervention: a nonlinear microstructural analysis	Stefan Reitz Mark P. Taylor
9	2006	Capital, labour and productivity: What role do they play in the potential GDP weakness of France, Germany and Italy?	Antonio Bassanetti Jörg Döpke, Roberto Torrini Roberta Zizza

10	2006	Real-time macroeconomic data and ex ante predictability of stock returns	J. Döpke, D. Hartmann C. Pierdzioch
11	2006	The role of real wage rigidity and labor market frictions for unemployment and inflation dynamics	Kai Christoffel Tobias Linzert
12	2006	Forecasting the price of crude oil via convenience yield predictions	Thomas A. Knetsch
13	2006	Foreign direct investment in the enlarged EU: do taxes matter and to what extent?	Guntram B. Wolff
14	2006	Inflation and relative price variability in the euro area: evidence from a panel threshold model	Dieter Nautz Juliane Scharff
15	2006	Internalization and internationalization under competing real options	Jan Hendrik Fisch
16	2006	Consumer price adjustment under the microscope: Germany in a period of low inflation	Johannes Hoffmann Jeong-Ryeol Kurz-Kim
17	2006	Identifying the role of labor markets for monetary policy in an estimated DSGE model	Kai Christoffel Keith Küster Tobias Linzert
18	2006	Do monetary indicators (still) predict euro area inflation?	Boris Hofmann
19	2006	Fool the markets? Creative accounting, fiscal transparency and sovereign risk premia	Kerstin Bernoth Guntram B. Wolff
20	2006	How would formula apportionment in the EU affect the distribution and the size of the corporate tax base? An analysis based on German multinationals	Clemens Fuest Thomas Hemmelgarn Fred Ramb

21	2006	Monetary and fiscal policy interactions in a New Keynesian model with capital accumulation and non-Ricardian consumers	Campbell Leith Leopold von Thadden
22	2006	Real-time forecasting and political stock market anomalies: evidence for the U.S.	Martin Bohl, Jörg Döpke Christian Pierdzioch
23	2006	A reappraisal of the evidence on PPP: a systematic investigation into MA roots in panel unit root tests and their implications	Christoph Fischer Daniel Porath
24	2006	Margins of multinational labor substitution	Sascha O. Becker Marc-Andreas Mündler
25	2006	Forecasting with panel data	Badi H. Baltagi
26	2006	Do actions speak louder than words? Household expectations of inflation based on micro consumption data	Atsushi Inoue Lutz Kilian Fatma Burcu Kiraz
27	2006	Learning, structural instability and present value calculations	H. Pesaran, D. Pettenuzzo A. Timmermann
28	2006	Empirical Bayesian density forecasting in Iowa and shrinkage for the Monte Carlo era	Kurt F. Lewis Charles H. Whiteman
29	2006	The within-distribution business cycle dynamics of German firms	Jörg Döpke Sebastian Weber
30	2006	Dependence on external finance: an inherent industry characteristic?	George M. von Furstenberg Ulf von Kalckreuth
31	2006	Comovements and heterogeneity in the euro area analyzed in a non-stationary dynamic factor model	Sandra Eickmeier

32	2006	Forecasting using a large number of predictors: is Bayesian regression a valid alternative to principal components?	Christine De Mol Domenico Giannone Lucrezia Reichlin
33	2006	Real-time forecasting of GDP based on a large factor model with monthly and quarterly data	Christian Schumacher Jörg Breitung
34	2006	Macroeconomic fluctuations and bank lending: evidence for Germany and the euro area	S. Eickmeier B. Hofmann, A. Worms
35	2006	Fiscal institutions, fiscal policy and sovereign risk premia	Mark Hallerberg Guntram B. Wolff
36	2006	Political risk and export promotion: evidence from Germany	C. Moser T. Nestmann, M. Wedow
37	2006	Has the export pricing behaviour of German enterprises changed? Empirical evidence from German sectoral export prices	Kerstin Stahn
38	2006	How to treat benchmark revisions? The case of German production and orders statistics	Thomas A. Knetsch Hans-Eggert Reimers
39	2006	How strong is the impact of exports and other demand components on German import demand? Evidence from euro-area and non-euro-area imports	Claudia Stirböck
40	2006	Does trade openness increase firm-level volatility?	C. M. Buch, J. Döpke H. Strotmann
41	2006	The macroeconomic effects of exogenous fiscal policy shocks in Germany: a disaggregated SVAR analysis	Kirsten H. Heppke-Falk Jörn Tenhofen Guntram B. Wolff

42	2006	How good are dynamic factor models at forecasting output and inflation? A meta-analytic approach	Sandra Eickmeier Christina Ziegler
43	2006	Regionalwährungen in Deutschland – Lokale Konkurrenz für den Euro?	Gerhard Rösl
44	2006	Precautionary saving and income uncertainty in Germany – new evidence from microdata	Nikolaus Bartzsch
45	2006	The role of technology in M&As: a firm-level comparison of cross-border and domestic deals	Rainer Frey Katrin Hussinger
46	2006	Price adjustment in German manufacturing: evidence from two merged surveys	Harald Stahl
47	2006	A new mixed multiplicative-additive model for seasonal adjustment	Stephanus Arz
48	2006	Industries and the bank lending effects of bank credit demand and monetary policy in Germany	Ivo J.M. Arnold Clemens J.M. Kool Katharina Raabe
01	2007	The effect of FDI on job separation	Sascha O. Becker Marc-Andreas Mündler
02	2007	Threshold dynamics of short-term interest rates: empirical evidence and implications for the term structure	Theofanis Archontakis Wolfgang Lemke
03	2007	Price setting in the euro area: some stylised facts from individual producer price data	Dias, Dossche, Gautier Hernando, Sabbatini Stahl, Vermeulen
04	2007	Unemployment and employment protection in a unionized economy with search frictions	Nikolai Stähler

05	2007	End-user order flow and exchange rate dynamics	S. Reitz, M. A. Schmidt M. P. Taylor
06	2007	Money-based interest rate rules: lessons from German data	C. Gerberding F. Seitz, A. Worms
07	2007	Moral hazard and bail-out in fiscal federations: evidence for the German Länder	Kirsten H. Heppke-Falk Guntram B. Wolff
08	2007	An assessment of the trends in international price competitiveness among EMU countries	Christoph Fischer
09	2007	Reconsidering the role of monetary indicators for euro area inflation from a Bayesian perspective using group inclusion probabilities	Michael Scharnagl Christian Schumacher
10	2007	A note on the coefficient of determination in regression models with infinite-variance variables	Jeong-Ryeol Kurz-Kim Mico Loretan

Series 2: Banking and Financial Studies

01	2006	Forecasting stock market volatility with macroeconomic variables in real time	J. Döpke, D. Hartmann C. Pierdzioch
02	2006	Finance and growth in a bank-based economy: is it quantity or quality that matters?	Michael Koetter Michael Wedow
03	2006	Measuring business sector concentration by an infection model	Klaus Düllmann
04	2006	Heterogeneity in lending and sectoral growth: evidence from German bank-level data	Claudia M. Buch Andrea Schertler Natalja von Westernhagen
05	2006	Does diversification improve the performance of German banks? Evidence from individual bank loan portfolios	Evelyn Hayden Daniel Porath Natalja von Westernhagen
06	2006	Banks' regulatory buffers, liquidity networks and monetary policy transmission	Christian Merkl Stéphanie Stolz
07	2006	Empirical risk analysis of pension insurance – the case of Germany	W. Gerke, F. Mager T. Reinschmidt C. Schmieder
08	2006	The stability of efficiency rankings when risk-preferences and objectives are different	Michael Koetter
09	2006	Sector concentration in loan portfolios and economic capital	Klaus Düllmann Nancy Masschelein
10	2006	The cost efficiency of German banks: a comparison of SFA and DEA	E. Fiorentino A. Karmann, M. Koetter
11	2006	Limits to international banking consolidation	F. Fecht, H. P. Grüner

12	2006	Money market derivatives and the allocation of liquidity risk in the banking sector	Falko Fecht Hendrik Hakenes
01	2007	Granularity adjustment for Basel II	Michael B. Gordy Eva Lütkebohmert
02	2007	Efficient, profitable and safe banking: an oxymoron? Evidence from a panel VAR approach	Michael Koetter Daniel Porath
03	2007	Slippery slopes of stress: ordered failure events in German banking	Thomas Kick Michael Koetter
04	2007	Open-end real estate funds in Germany – genesis and crisis	C. E. Bannier F. Fecht, M. Tyrell
05	2007	Diversification and the banks' risk-return-characteristics – evidence from loan portfolios of German banks	A. Behr, A. Kamp C. Memmel, A. Pfingsten
06	2007	How do banks adjust their capital ratios? Evidence from Germany	Christoph Memmel Peter Raupach

Visiting researcher at the Deutsche Bundesbank

The Deutsche Bundesbank in Frankfurt is looking for a visiting researcher. Among others under certain conditions visiting researchers have access to a wide range of data in the Bundesbank. They include micro data on firms and banks not available in the public. Visitors should prepare a research project during their stay at the Bundesbank. Candidates must hold a Ph D and be engaged in the field of either macroeconomics and monetary economics, financial markets or international economics. Proposed research projects should be from these fields. The visiting term will be from 3 to 6 months. Salary is commensurate with experience.

Applicants are requested to send a CV, copies of recent papers, letters of reference and a proposal for a research project to:

Deutsche Bundesbank
Personalabteilung
Wilhelm-Epstein-Str. 14

60431 Frankfurt
GERMANY