DISSERTATION

ESTIMATION AND LINEAR PREDICTION FOR REGRESSION, AUTOREGRESSION AND ARMA WITH INFINITE VARIANCE DATA

Submitted by

Daren B. H. Cline

Department of Statistics

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 1983

COLORADO STATE UNIVERSITY

July 7, 1983

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY DAREN B. H. CLINE ENTITLED ESTIMATION AND LINEAR PREDICTION FOR REGRESSION, AUTOREGRESSION AND ARMA WITH INFINITE VARIANCE DATA BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Richard Davis

Duane C Boes

John Locker

Adviser Sidney Resnick

Duane C Boes
Department Head

ABSTRACT OF DISSERTATION

ESTIMATION AND LINEAR PREDICTION FOR REGRESSION, AUTOREGRESSION AND ARMA WITH INFINITE VARIANCE DATA

This dissertation is divided into four parts, each of which considers random variables from distributions with regularly varying tails and/or in a stable domain of attraction. Part I considers the existence of infinite series of an independent sequence of such random variables and the relationship of the probability of large values of the series to the probability of large values of the first component. Part II applies Part I in order to provide a linear predictor for ARMA time series (again with regularly varying tails). This predictor is designed to minimize the probability of large prediction errors relative to the tails of the noise distribution. Part III investigates the products of independent random variables where one has distribution in a stable domain of attraction and gives conditions for which the product distribution is in a stable domain of attraction. Part IV consideres estimation of the regression parameter in a model where the independent variables are in a stable domain of attraction. Consistency for certain M-estimators is proved. Utilizing portions of Part III this final part gives necessary and sufficient conditions

for consistency of least squares estimators and provides the asymptotic distribution of least squares estimators.

Daren B. H. Cline
Statistics Department
Colorado State University
Fort Collins, Colorado 80523
Summer 1983

ACKNOWLEDGEMENTS

My greatest thanks go to my two advisers, Professors Peter J. Brockwell and Sidney I. Resnick: to Pete who led me into the fascinating world of stable laws and who gently persuaded me to look at these problems, to Sid who frequently saved me with his intimate knowledge of regular variation and who kept me going in the final weeks, and to both who patiently reviewed my too often faulty work.

I would also like to thank the remainder of my committee, Professors Richard A. Davis, Duane C. Boes and John Locker, for their generous encouragement and praise. Special thanks go to Richard who also helped in the editing.

To my fellow classmates, I express fond appreciation for all the good times together, in spite of the rigors of graduate school. To Mary Frary, I give tremendous thanks for her expert typing with no time to spare.

And finally, to my wife, Marlene Hsi, I offer the deepest love and gratitude for her tender support, coming as it did at the time of her own rigorous doctoral research.

TABLE OF CONTENTS

	Page
OVERVIEW	1
PART I: INFINITE SERIES OF RANDOM VARIABLES WITH REGULARLY VARYING TAILS	3
1. Introduction	4
2. Existence and Tail Behavior of Infinite Series	8
3. Dispersion as a Metric	18
4. ARMA Processes With Regularly Varying Tails	27
PART II: LINEAR PREDICTION OF ARMA PROCESSES WITH INFINITE VARIANCE	33
1. Introduction	34
2. Minimum Dispersion Prediction for Autoregressive Processes	39
3. Prediction of the ARMA(1,1) Process	42
4. Prediction for the MA(q) and ARMA(p,q) Models	52
5. Numerical Comparison of Minimum Dispersion and Least Squares Predictors	57
PART III: PRODUCTS OF INDEPENDENT RANDOM VARIABLES AND DOMAINS OF ATTRACTION	68
1. Introduction	69
2. Regular Variation of the Product Distribution	71
3. Stable Attraction of the Product Distribution	78
4. Joint Attraction of Two Products	87

Page

PART IV: REGRESSION WITH INFINITE VARIANCE DATA 100

1. Introduction 101

2. Consistency of M-Estimators 105

3. Asymptotic Distribution of Least Squares Estimators 122

REFERENCES 127

LIST OF TABLES

Table		Page
5.1	A Comparison of Error Dispersion in Predicting X_{n+1} from an ARMA(1,1) Process	60
5.2	2500 Series of Size 10, Predicting Observation 11: $\alpha = 0.50 \quad \phi = .000 \quad \theta = .800$	62
5.3	2500 Series of Size 10, Predicting Observation 11: $\alpha = 1.00 \quad \phi = .700 \quad \theta = .800$	64
5.4	1000 Series of Size 25, Predicting Observation 26: $\alpha = 0.50 \quad \phi = .700 \quad \theta = .800$	66

OVERVIEW

The common theme in the four parts of this dissertation is the application to statistical inference of the theory of distributions with regularly varying tails. Parts I and III lay the groundwork for the specific applications in Parts II and IV and though they are written with the applications in mind, they consider problems which are of independent interest. Our ultimate objective was to study both estimation and linear prediction for autoregressive-moving average (ARMA) time series with regularly varying tails and as a stepping stone, estimation for regression. Time, however, has required us to consider estimation for the regression model only.

Part I studies infinite series of the form $Y = \sum_{j=1}^{\infty} \rho_j W_j$ where the $\{W_j\}$ is an iid sequence of random variables with regularly varying tails. We consider conditions for the existence of Y and relate the probability of large values of Y to the tail of W_1. This relationship turns out to be similar to a common metric for ℓ_α sequence spaces. Capitalizing on this, we define projection operators for the linear space generated by $\{W_j\}$. Part I also shows the ARMA time series can be expressed as infinite order moving average sequences.

The results of the first part are applied in Part II to ARMA processes with regularly varying tails. Using the metric defined in Part I, we consider linear prediction by minimizing the probability of large prediction errors. Prediction for the AR(ρ) and ARMA (1,1) models is investigated thoroughly. The procedure is similar to least squares prediction and comparisons are made.

Part III presents theory used partially in the final part. In Part III we consider conditions for which the distribution of the product of independent random variables is in a stable domain of attraction when at least one has distribution in a stable domain of attraction. The theory is extended to include bivariate distributions of the component-wise product of two independent pairs of random variables.

Estimation of the regression parameter appears in Part IV. Our regression model assumes the independent variable is in a stable domain of attraction. Using results in the literature, we consider consistency of M-estimators. Least squares estimation, however, is given the fullest treatment and we provide necessary and sufficient conditions for the least squares estimator to be weakly consistent. Limit distributions are described for the least squares estimator and we demonstrate the startling result that the limit can be either normal or the ratio of two non-normal stables, depending on the distribution of the independent variable.

PART I: INFINITE SERIES OF RANDOM VARIABLES WITH REGULARLY VARYING TAILS

Summary. We give conditions for the convergence of an infinite series of independent and identically distributed random variables, whose distribution has regularly varying tails. More importantly, we show that the distribution of such a series is tail equivalent to the distribution of its components. This enables us to define a quantity, which we call dispersion, measuring the relative thickness of the tails and thereby to compare different infinite series. The dispersion may be related to the ℓ_α metric for sequence spaces and this leads to a notion of linear projection which is useful for prediction of time series. Since time series prediction is our ultimate objective, we also discuss the notions of stationarity, causality and invertibility for ARMA processes which are driven by random variables with regularly varying tails.

1. Introduction

We are concerned with random variables of the form $Y = \sum_{j=1}^{\infty} \rho_j W_j$ where $\{W_j\}$ is a sequence of independent random variables, all from a distribution F^* with regularly varying tails. Let F be the distribution of $|W_j|$ and define $\bar{F}(t) = 1-F(t) = P[|W_j|>t]$. We say that $\bar{F}(t)$ is regularly varying with exponent $-\alpha$ ($\bar{F}\varepsilon RV_{-\alpha}$) if for every $s>0$

$$\lim_{t\to\infty} \frac{\bar{F}(st)}{\bar{F}(t)} = s^{-\alpha}.$$

In fact, if the limit exists at all, then it will be of the form $s^{-\alpha}$ for some $\alpha>0$ and the convergence will be uniform on $[s_0,\infty)$ for any $s_0>0$. Furthermore, for any $\varepsilon>0$ there exists $c>0$ such that $\bar{F}(t)\leq ct^{-\alpha+\varepsilon}$ for all $t\geq s_0$. (See Feller II for a discussion of more general regularly varying functions.) The parameter α we call the <u>tail index</u> of F. Distributions with index α have moments up to (and perhaps including) order α, but higher moments do not exist. In particular, if $\alpha<2$ then the variance does not exist. More precisely, Feller (1971), p. 283, proves the following relationship between F and its truncated moments.

<u>Lemma 1.1</u> Suppose $|W|$ has distribution F^* with $\bar{F} = P[|W|>t] \varepsilon RV_{-\alpha}$. Define $m_\gamma(t) = E[|W|^\gamma 1_{|W|\leq t}]$ and, when it exists, $u_\gamma(t) = E[|W|^\gamma 1_{|W|>t}]$. Then for $\gamma\geq\alpha$, $t^{-\gamma}m_\gamma(t) \varepsilon RV_{-\alpha}$ and $\lim_{t\to\infty} \frac{m_\gamma(t)}{t^\gamma\bar{F}(t)} = \frac{\alpha}{\gamma-\alpha}$, and when $u_\gamma(t)<\infty$, $t^{-\gamma}u_\gamma(t)\varepsilon RV_{-\alpha}$ and $\lim_{t\to\infty} \frac{u_\gamma(t)}{t^\gamma\bar{F}(t)} = \frac{\alpha}{\alpha-\gamma}$. #

Random variables with regularly varying tails exhibit a striking relationship between the distributions of sums and of maxima. The following is a modification of a theorem in Feller (1971), p. 278.

<u>Lemma 1.2</u> Suppose $W_1, W_2, \ldots, W_n$ are independent and identically distributed. Let F be the distribution of $|W_j|$ and suppose $\bar{F} \varepsilon RV_{-\alpha}$. For real numbers $\rho_1, \ldots, \rho_n$, define $\bar{G}(t) = P\left[\left|\sum_{j=1}^{n} \rho_j W_j\right| > t\right]$ and $\bar{H}(t) = P\left[\sup_{1 \le j \le n} |\rho_j W_j| > t\right]$. Then $\bar{G} \varepsilon RV_{-\alpha}$, $\bar{H} \varepsilon RV_{-\alpha}$ and

$$\lim_{t\to\infty} \frac{\bar{G}(t)}{\bar{F}(t)} = \lim_{t\to\infty} \frac{\bar{H}(t)}{\bar{F}(t)} = \sum_{j=1}^{n} |\rho_j|^{\alpha}.$$

<u>Proof</u>: The result, if true for n=2, extends by induction. We therefore consider only the case n=2. First,

$$\lim_{t\to\infty} \frac{\bar{H}(t)}{\bar{F}(t)} = \lim_{t\to\infty} \left(\frac{\bar{F}(t/|\rho_1|) + \bar{F}(t/|\rho_2|) - \bar{F}(t/|\rho_1|)\bar{F}(t/|\rho_2|)}{\bar{F}(t)} \right)$$

$$= |\rho_1|^{\alpha} + |\rho_2|^{\alpha} + 0$$

by the regular variation principle.

Second, by an application of the theorem in Feller (1971), p. 278,

$$\overline{\lim_{t\to\infty}} \frac{\bar{G}(t)}{\bar{F}(t)} \le \lim_{t\to\infty} \frac{P[|\rho_1 W_1| + |\rho_2 W_2| > t]}{\bar{F}(t)}$$

$$= \lim_{t\to\infty} \frac{\bar{F}(t/|\rho_1|) + \bar{F}(t/|\rho_2|)}{\bar{F}(t)}$$

$$= |\rho_1|^{\alpha} + |\rho_2|^{\alpha}. \qquad (1.1)$$

However, for any $\delta > 0$

$$\begin{aligned}\bar{G}(t) &= P[|\rho_1 W_1 + \rho_2 W_2| > t] \\ &\geq P[|\rho_1 W_1| > (1+\delta)t, |\rho_2 W_2| \leq \delta t] + P[|\rho_2 W_2| > (1+\delta)t, |\rho_1 W_1| \leq \delta t] \\ &= \bar{F}\left(\frac{(1+\delta)t}{|\rho_1|}\right) F\left(\frac{\delta t}{|\rho_2|}\right) + \bar{F}\left(\frac{(1+\delta)t}{|\rho_2|}\right) \bar{F}\left(\frac{\delta t}{|\rho_1|}\right)\end{aligned}$$

And from this we calculate

$$\underline{\lim}_{t\to\infty} \frac{\bar{G}(t)}{\bar{F}(t)} \geq (1+\delta)^{-\alpha}|\rho_1|^{\alpha} + (1+\delta)^{-\alpha}|\rho_2|^{\alpha}. \tag{1.2}$$

Since δ is arbitrary, then (1.2) combined with (1.1) gives us the result. #

Lemma 1.2 tells us in particular, how to compare the tail of the distribution of a sum with the tail of the distribution of each component. We will extend this result to infinite series in Section 2. Whenever two distributions with regularly varying tails (say F_1 and F_2) satisfy $\lim_{t\to\infty} \frac{\bar{F}_1(t)}{\bar{F}_2(t)}$ exists and is nonzero, we say that F_1 and F_2 are tail equivalent. The limiting ratio gives us a convenient means to compare the probability of large values of random variables from the two distributions. In particular, we may be interested in the probability that $|Y_1| = |\sum_{j=1}^{\infty} \rho_{1j} W_j|$ is large relative to the probability that $|Y_2| = |\sum_{j=1}^{\infty} \rho_{2j} W_j|$ is large. For example, Y_1 and Y_2 might be the prediction errors from alternate methods of predicting a time series and we may prefer to choose the predictor which has the least chance of large errors (see Part II).

The limiting ratio of probabilities for $Y = \sum_{j=1}^{\infty} \rho_j W_j$ and W_1 will turn out to be $\sum_{j=1}^{\infty} |\rho_j|^{\alpha}$, a quantity we will call the <u>dispersion of Y</u>. When comparing variables on the linear space generated by a given sequence $\{W_j\}$, the dispersion is a useful measure of distance. This leads to the concept of minimum dispersion projection for variables in this linear space. Section 3 investigates this notion.

In Section 4, we discuss the existence of a stationary ARMA time series driven by regular varying tail noise and conditions for causality and invertibility of such a time series.

2. Existence and Tail Behavior of Infinite Series

We start with an application of Kolmogorov's three series theorem to series of regularly varying tail variables.

Theorem 2.1 Suppose $\{W_j\}$ are iid F^* and $\bar{F}(t) = P[|W_j|>t]\varepsilon RV_{-\alpha}$.

Then $Y = \lim_{n\to\infty} \sum_{j=1}^{n} \rho_j W_j$ exists almost surely if either

i) $\sum_{j=1}^{\infty} |\rho_j|^{\delta}<\infty$ for some $\delta<\alpha$, $\delta\leq 1$

or

ii) EW_j exists and equals 0, and $\sum_{j=1}^{\infty} |\rho_j|^{\delta} < \infty$

for some $\delta<\alpha, \delta\leq 2$ (or $\delta=1$ if $\alpha=1$).

Proof:

i) The series $Y = \sum_{j=1}^{\infty} \rho_j W_j$ is absolutely convergent if and only if for all $v>0$

$$\sum_{j=1}^{\infty} P[|\rho_j W_j|>v] = \sum_{j=1}^{\infty} \bar{F}(v/|\rho_j|) < \infty$$

and
$$\sum_{j=1}^{\infty} E\left[|\rho_j W_j| 1_{|\rho_j W_j|\leq v}\right] = \sum_{j=1}^{\infty} |\rho_j| m_1(v/|\rho_j|) < \infty.$$

(The third series is not necessary to prove absolute convergence.)

Since by Lemma 2.1,

$$\lim_{t\to\infty} \frac{\bar{F}(t)}{t^{-1}m_1(t)} = \begin{cases} \frac{1-\alpha}{\alpha} & \text{if } \alpha\leq 1 \\ 0 & \text{otherwise} \end{cases},$$

it suffices to show the second series converges.

If $\alpha\leq 1$ then $t^{-1}m_1(t)\varepsilon RV_{-\alpha}$ and so there exists a $c>0$ such that for any $s>\frac{v}{\sup_j|\rho_j|}$, $s^{-1}m_1(s)\leq cs^{-\delta}$, if $\delta<\alpha$. If $\alpha>1$ then $m_1(t)\to E|W_j|$ and so we can use the bound $s^{-1}m_1(s)\leq cs^{-\delta}$ if $\delta\leq 1$. In either case,

$$\sum_{j=1}^{\infty} |\rho_j| m_1(v/|\rho_j|) \leq cv^{-\delta}\sum_{j=1}^{\infty} |\rho_j|^{\delta} < \infty.$$

Thus, condition i) is sufficient for absolute convergence of Y.

ii) If $EW_j=0$ (in which case $\alpha\geq 1$), then

$$\begin{aligned} \left|E\left[W_j 1_{|W_j|\leq t}\right]\right| &= \left|E\left[-W_j 1_{|W_1|>t}\right]\right| \\ &\leq E\left[|W_j| 1_{|W_j|>t}\right] \\ &= u_1(t) \end{aligned}$$

For Y to exist, it suffices to prove that

$$\sum_{j=1}^{\infty} P[|\rho_j W_j|>v] = \sum_{j=1}^{\infty} \bar{F}(v/|\rho_j|) < \infty$$

$$\sum_{j=1}^{\infty} \left|\rho_j E\left[W_j 1_{|\rho_j W_j|\leq v}\right]\right| \leq \sum_{j=1}^{\infty} |\rho_j| u_1(v/|\rho_j|) < \infty$$

and

$$\sum_{j=1}^{\infty} E\left[(\rho_j W_j)^2 1_{|\rho_j W_j|\leq v}\right] = \sum_{j=1}^{\infty} \rho_j^2 m_2(v/|\rho_j|) < \infty.$$

From Lemma 2.1

$$\lim_{t\to\infty} \frac{\bar{F}(t)}{t^{-1}u_1(t)} = \frac{\alpha-1}{\alpha}$$

$$\lim_{t\to\infty} \frac{\bar{F}(t)}{t^{-2}m_2(t)} = \begin{cases} \frac{2-\alpha}{\alpha} & \text{if } \alpha\leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Thus if $\alpha=1$, convergence of the second series is sufficient and if $\alpha>1$, convergence of the third is sufficient. For $\alpha=1$, since $u_1(t)\to 0$ there exists c such that for all $s\geq \frac{v}{\sup_j|\rho_j|}$, $s^{-1}u_1(s) \leq cs^{-1}$ so that

$$\sum_{j=1}^{\infty} |\rho_j|u_1(v/|\rho_j|) \leq c\sum_{j=1}^{\infty} |\rho_j|$$

and hence condition ii) guarantees that Y exists.

For $\alpha>1$, we can find c such that $s^{-2}m_2(s)\leq cs^{-\delta}$ where $\delta<\alpha$, $\delta\leq 2$, so that

$$\sum_{j=1}^{\infty} |\rho_{j_1}|^2 m_2(v/|\rho_j|) \leq cu^{2-\delta}\sum_{j=1}^{\infty} |\rho_j|^{\delta}$$

and again condition ii) is sufficient. #

We remark that when $\sum_{j=1}^{\infty} \rho_j W_j$ is absolutely convergent then $\sup_j|\rho_j W_j|$ exists almost surely, also.

Sometimes the condition $\sum_{j=1}^{\infty} |\rho_j|^{\alpha} < \infty$ is sufficient for the existence of $\sum_{j=1}^{\infty} \rho_j W_j$. For an example, assume the W_j are symmetric about 0 and $\bar{F}\varepsilon RV_{-\alpha}$, $0<\alpha<2$. In this case it is sufficient to show

$$\sum_{j=1}^{\infty} F(v/|\rho_j|) < \infty \text{ for all } u>0.$$

If $\bar{F}$ satisfies $\overline{\lim_{t\to\infty}}\, t^{\alpha}\bar{F}(t) < \infty$, then there exists c such that for

$s > \dfrac{u}{\sup_j|\rho_j|}$, $\bar{F}(t) \le cs^{-\alpha}$. Thus

$$\sum_{j=1}^{\infty} \bar{F}(v/|\rho_j|) \le cv^{-\alpha}\sum_{j=1}^{\infty}|\rho_j|^{\alpha} < \infty$$

and hence $\sum_{j=1}^{\infty}\rho_j W_j$ exists almost surely.

On the other hand, a counterexample is the following. Suppose $\{W_j\}$ are distributed so that for t large enough, $P[|W_j|>t] = \bar{F}(t)$ $= t^{-\alpha}\ln t$. Suppose also that $\rho_j = (j(\ln j)^2)^{-1/\alpha}$; $j\geq 2$. Then $\sum_{j=2}^{\infty}\rho_j^{\alpha} < \infty$, but with j_0 chosen large enough,

$$\sum_{j=j_0}^{\infty} \bar{F}(1/\rho_j) = \frac{1}{\alpha}\sum_{j=j_0}^{\infty}\frac{\ln(j(\ln j)^2)}{j(\ln j)^2}$$

$$\geq \frac{1}{\alpha}\sum_{j=j_0}^{\infty}\frac{1}{j\ln j} = \infty.$$

Therefore $\sum_{j=2}^{\infty}\rho_j W_j$ almost surely does not exist.

<u>Lemma 2.3</u> Suppose F* is a probability measure for W with F the distribution of $|W|$ and $\bar{F}(t) = P[|W|>t]\varepsilon RV_{-\alpha}, \alpha > 0$. Suppose also that $\{\rho_j\}$ satisfies $\sum_{j=1}^{\infty}|\rho_j|^{\delta} < \infty$ for some $\delta<\alpha$. Then

$$\text{i)}\quad \lim_{t\to\infty}\frac{\sum_{j=1}^{\infty}\bar{F}(t/|\rho_j|)}{\bar{F}(t)} = \sum_{j=1}^{\infty}|\rho_j|^{\alpha}$$

and

$$\text{ii)}\quad \lim_{t\to\infty}\frac{1-\prod_{j=1}^{\infty}F(t/|\rho_j|)}{\bar{F}(t)} = \sum_{j=1}^{\infty}|\rho_j|^{\alpha}.$$

Proof:

i) Let $m = \sup_j |\rho_j|$. There exists $c>0$, $t_0>0$ such that for all $t>t_0$, $y > 1/m$.

$$\frac{\bar{F}(ty)}{\bar{F}(t)} \le cy^{-\delta}.$$

Therefore $\frac{1}{\bar{F}(t)} \sum_{j=1}^{\infty} \bar{F}(t/|\rho_j|) \le \sum_{j=1}^{\infty} |\rho_j|^{\delta} < \infty$.

Since $$\lim_{t\to\infty} \frac{\bar{F}(t/|\rho_j|)}{\bar{F}(t)} = |\rho_j|^{\alpha}$$

Then by dominated convergence the result i) holds.

ii) By i) $\sum_{j=1}^{\infty} \bar{F}(t/|\rho_j|) < \infty$ for all $t>0$ and $\sup_j \bar{F}(t/|\rho_j|)\to 0$ as $t\to\infty$.

We can therefore exchange $\bar{F}$ with $\ln F$ to get

$$\lim_{t\to\infty} \frac{\sum_{j=1}^{\infty} \bar{F}(t/|\rho_j|)}{1 - \prod_{j=1}^{\infty} F(t/|\rho_j|)} = \lim_{t\to\infty} \frac{\sum_{j=1}^{\infty} \bar{F}(t/|\rho_j|)}{\sum_{j=1}^{\infty} \ln F(t/|\rho_j|)} = 1$$

With i), this implies ii). #

The main result of this section is next.

Theorem 2.4 Suppose $\{W_j\} \sim$ iid F^* where $\bar{F}(t) = P[|W_j|>t] \in RV_{-\alpha}$ and suppose $\{\rho_j\}$ satisfy $\sum_{j=1}^{\infty} |\rho_j|^{\delta} < \infty$ for some $\delta<\alpha, \alpha\le 1$. Let $\bar{G}(t) = P[|\sum_{j=1}^{\infty} \rho_j W_j|>t]$ and $\bar{H}(t) = P[\sup_j |\rho_j W_j|>t]$. Then $\lim_{t\to\infty} \frac{\bar{G}(t)}{\bar{F}(t)} = \lim_{t\to\infty} \frac{\bar{H}(t)}{\bar{F}(t)} = \sum_{j=1}^{\infty} |\rho_j|^{\alpha}$.

Proof: Set $Y = \sum_{j=1}^{\infty} \rho_j W_j$, $Y_n = \sum_{j=1}^{n} \rho_j W_j$, $Z_1 = \sum_{j=1}^{\infty} |\rho_j W_j|$ and $Z_{\infty} = \sup_j |\rho_j W_j|$.

By Theorem 2.1, Z_1 exists almost surely and hence Z_{∞} and Y do also.

Since $H(t) = \prod_{j=1}^{\infty} (1-\bar{F}(t/|\rho_j|)$, Lemma 2.3 immediately gives the second conclusion.

To prove the first, we let G_n be the distribution of Y_n. Then for any $n \geq 1$, $\varepsilon > 0$,

$$P[|Y|>t] \geq P[|Y_n|>(1+\varepsilon)t,\ |Y-Y_n|<\varepsilon t]$$

$$= \bar{G}_n((1+\varepsilon)t)P[|Y-Y_n|<\varepsilon t].$$

Thus,

$$\varliminf_{t\to\infty} \frac{\bar{G}(t)}{\bar{F}(t)} \geq \lim_{t\to\infty} \frac{\bar{G}_n((1+\varepsilon)t)}{\bar{F}(t)} P[|Y-Y_n|<\varepsilon t]$$

$$= (1+\varepsilon)^{-\alpha} \sum_{j=1}^{n} |\rho_j|^{\alpha},$$

where the limit is obtained by using Lemma 1.2 and the fact that $\bar{F}\varepsilon RV_{-\alpha}$. Since both n and ε are arbitrary,

$$\varliminf_{t\to\infty} \frac{\bar{G}(t)}{\bar{F}(t)} \geq \sum_{j=1}^{\infty} |\rho_j|^{\alpha} \tag{2.1}$$

The alternate inequality is first proven for $\alpha<1$. Define $\phi(\lambda) = Ee^{-\lambda|W_j|}$. When $\alpha<1$, then by Feller (1971), p. 447,

$$\lim_{t\to\infty} \frac{1-\phi(1/t)}{\bar{F}(T)} = \Gamma(1+\alpha) \tag{2.2}$$

This indicates that $1-\phi(1/t)$ is a regularly varying distribution tail, so that Lemma 2.2 applies.

$$\lim_{t\to\infty} \frac{1-\prod_{j=1}^{\infty} \phi(|\rho_j|/t)}{1-\phi(1/t)} = \sum_{j=1}^{\infty} |\rho_j|^{\alpha} \tag{2.3}$$

Of course, $Ee^{-\lambda Z_1} = \prod_{j=1}^{\infty} \phi(\lambda|\rho_j|)$. Let H_1 be the distribution of Z_1. Applying the theorem in Feller again, $\bar{H}_1 \in RV_{-\alpha}$ and

$$\lim_{t\to\infty} \frac{1 - \prod_{j=1}^{\infty} \phi(|\rho_j|/t)}{H_1(t)} = \Gamma(1+\alpha) \tag{2.4}$$

Combining (2.2), (2.3) and (2.4) we have

$$\lim_{t\to\infty} \frac{\bar{H}_1(t)}{\bar{F}(t)} = \sum_{j=1}^{\infty} |\rho_j|^{\alpha} .$$

Since $|Y| \leq Z_1$, then

$$\overline{\lim_{t\to\infty}} \frac{\bar{G}(t)}{\bar{F}(t)} \leq \lim_{t\to\infty} \frac{\bar{H}_1(t)}{\bar{F}(t)} = \sum_{j=1}^{\infty} |\rho_j|^{\alpha}$$

which, together with (2.1), proves the result for $\alpha<1$.

When $\alpha \geq 1$, then $a = \sum_{j=1}^{\infty} |\rho_j| < \infty$. Let $\gamma > \alpha$ and $p_j = \frac{1}{a}|\rho_j|$. By Holder's inequality, with $\{p_j\}$ as the probability measure,

$$\begin{aligned} Z_1 &= a \sum_{j=1}^{\infty} |W_j| p_j \\ &\leq a\left(\sum_{j-1}^{\infty} |W_j|^{\gamma} p_j\right)^{1/\gamma} \\ &= a^{1-1/\gamma}\left(\sum_{j=1}^{\infty} |W_j|^{\gamma}|\rho_j|\right)^{1/\gamma} \end{aligned} \tag{2.5}$$

The distribution of $|W_j|^{\gamma}$ is $F(t^{1/\gamma})$ and has index $\alpha/\gamma<1$. Letting $V = \sum_{j=1}^{\infty} |W_j|^{\gamma}|\rho_j|$ and relying on the result for index less than 1,

$$\lim_{t\to\infty} \frac{P[V>t]}{\bar{F}(t^{1/\gamma})} = \sum_{j=1}^{\infty} |\rho_j|^{\alpha/\gamma}. \tag{2.6}$$

From (2.5) and (2.6) we can calculate,

$$\begin{aligned}
\overline{\lim_{t\to\infty}} \frac{\bar{H}_1(t)}{\bar{F}(t)} &\le \lim_{t\to\infty} \frac{P[a^{1-1/\gamma}V^{1/\gamma} > t]}{\bar{F}(t)} \\
&= \lim_{s\to\infty} \frac{P[V > a^{1-\gamma}s]}{F(s^{1/\gamma})} \\
&= (a^{1-\gamma})^{-\alpha/\gamma} \sum_{j=1}^{\infty} |\rho_j|^{\alpha/\gamma} \\
&= \left(\sum_{j=1}^{\infty} |\rho_j| \right)^{\alpha-\alpha/\gamma} \sum_{j=1}^{\infty} |\rho_j|^{\alpha/\gamma}.
\end{aligned}$$

Since $\gamma > \alpha$ is arbitrary,

$$\overline{\lim_{t\to\infty}} \frac{\bar{H}_1(t)}{\bar{F}(t)} \le \left(\sum_{j=1}^{\infty} |\rho_j| \right)^{\alpha}. \tag{2.7}$$

This is still not strong enough to prove our result. However, with $Y_n = \sum_{j=1}^{n} \rho_j W_j$ and $\varepsilon < \frac{1}{2}$,

$$\begin{aligned}
P[|Y| > t] &\le P[|Y_n| > (1-\varepsilon)t] + P[|Y-Y_n| > (1-\varepsilon)t] \\
&\quad + P[|Y_n| > \varepsilon t,\ |Y-Y_n| > \varepsilon t]. \\
&= \bar{G}_n((1-\varepsilon)t) + \bar{G}_{-n}((1-\varepsilon)t) + \bar{G}_n(\varepsilon t)\bar{G}_{-n}(\varepsilon t)
\end{aligned} \tag{2.8}$$

where G_{-n} is the distribution of $Y-Y_n$, which is independent of Y_n. By Lemma 1.2 and the inequality (2.7), respectively,

$$\lim_{t\to\infty} \frac{\bar{G}_n(t)}{\bar{F}(t)} = \sum_{j=1}^{n} |\rho_j|^\alpha$$

and

$$\overline{\lim_{t\to\infty}} \frac{\bar{G}_{-n}(t)}{\bar{F}(t)} \le \overline{\lim_{t\to\infty}} \frac{P\left[\sum_{j=n+1}^{\infty} |\rho_j W_j| > t\right]}{\bar{F}(t)}$$

$$\le \left(\sum_{j=n+1}^{\infty} |\rho_j|\right)^\alpha.$$

Using these in (2.8),

$$\overline{\lim_{t\to\infty}} \frac{\bar{G}(t)}{\bar{F}(t)} \le (1-\varepsilon)^{-\alpha} \sum_{j=1}^{n} |\rho_j|^\alpha + (1-\varepsilon)^{-\alpha}\left(\sum_{j=n+1}^{\infty} |\rho_j\right)^\alpha.$$

Since n and ε are arbitrary, then

$$\overline{\lim_{t\to\infty}} \frac{\bar{G}(t)}{\bar{F}(t)} \le \sum_{j=1}^{\infty} |\rho_j|^\alpha$$

and with (2.1) we have our result for $\alpha \ge 1$. #

The quantity $\sum_{j=1}^{\infty} |\rho_j|^\alpha$ we call the dispersion of Y (disp(Y)). This theorem indicates that disp(Y) is a measure of the probability of large values of Y. If $\{W_j\}$ are symmetric stable (α) in distribution, then Y will also be symmetric stable (α) and $(\mathrm{disp}(Y))^{1/\alpha}$ will be the ratio of Y's scale parameter to W_j's scale. Section 3 demonstrates how dispersion may be used as a measure of distance between random variables which are infinite series in $\{W_j\}$.

Corollary 2.5 Let $Z_\gamma = \left[\sum_{j=1}^{\infty} |\rho_j W_j|^\gamma\right]^{1/\gamma}$ for $\gamma \geq \delta$, then

$$\lim_{t\to\infty} \frac{P[Z_\gamma > t]}{P[|W|>y]} = \sum_{j=1} |\rho_j|^\alpha .$$

Proof: $|W_j|^\gamma$ has distribution tail $\bar{F}_\gamma(t) = \bar{F}(t^{1/\gamma}) \varepsilon RV_{-\alpha/\gamma}$. For $\delta_1 = \delta/\gamma$, $\delta_1 \leq 1$ and $\delta_1 < \alpha/\gamma$ and $\sum_{j=1}^{\infty} \left(|\rho_j|^\gamma\right)^{\delta_1} < \infty$. Thus Z_γ exists almost surely. Let H_γ be the distribution of $(Z_\gamma)^\gamma = \sum_{j=1}^{\infty} |\rho_j W_j|^\gamma$. By the theorem,

$$\lim_{t\to\infty} \frac{\bar{H}_\gamma(t)}{\bar{F}_\gamma(t)} = \sum_{j-1}^{\infty} \left(|\rho_j|^\gamma\right)^{\alpha/\gamma} = \sum_{j=1}^{\infty} |\rho_j|^\alpha .$$

Thus

$$\lim_{t\to\infty} \frac{P[Z_\gamma > t]}{P[|W|>t]} = \lim_{t\to\infty} \frac{\bar{H}_\gamma(t^{1/\gamma})}{\bar{F}(t^{1/\gamma})} = \sum_{j=1}^{\infty} |\rho_j|^\alpha . \qquad \#$$

3. Dispersion as a Metric

In this section we define a metric for infinite series of regularly varying variables and a corresponding projection operator. We also elaborate on the nature of the projection operator. As before, the sequence $\{W_j\}$ will be independent and identically distributed, $\bar{F}(t) = P[|W_j|>t]$ is the tail of $|W_j|$ and regularly varying with exponent $-\alpha$. Recall that we have defined the dispersion of $Y = \sum_{j=1}^{\infty} \rho_j W_j$ by

$$\operatorname{disp}(Y) = \sum_{j=1}^{\infty} |\rho_j|^{\alpha} .$$

Let $\delta>0$ satisfy $\delta<\alpha, \alpha\leq 1$. Define now the (random) linear space for given sequence $\{W_j\}$,

$$\mathcal{S} = \{Y = \sum_{j=1}^{\infty} \rho_j W_j \text{ such that } \sum_{j=1}^{\infty} |\rho_j|^{\delta} < \infty\}.$$

We remark that in fact we need only work with a space of equivalence classes which are well defined by the distribution structure, but $\mathcal{S}$ is a convenient means to express this. For $Y_1 = \sum_{j=1}^{\infty} \rho_{1j} W_j$, $Y_2 = \sum_{j=1}^{\infty} \rho_{2j} W_j$ we define

$$d(Y_1, Y_2) = \begin{cases} \sum_{j=1}^{\infty} |\rho_{1j}-\rho_{2j}|^{\alpha} = \operatorname{disp}(Y_1-Y_2) & \text{if } \alpha \leq 1 \\ \left(\sum_{j=1}^{\infty} |\rho_{1j}-\rho_{2j}|^{\alpha}\right)^{1/\alpha} = (\operatorname{disp}(Y_1-Y_2))^{1/\alpha} & \text{if } \alpha > 1. \end{cases}$$

Lemma 3.1 d is a metric on S.

Proof: The only condition not obvious is that $d(Y_1,Y_2) = 0$ if and only if $Y_1=Y_2$ almost surely. Clearly, if $d(Y_1,Y_2) = 0$ then $\rho_{1j}=\rho_{2j}$ for all j and hence

$$\begin{aligned} Y_1-Y_2 &= \lim_{n\to\infty}\left(\sum_{j=1}^{n} \rho_{1j}W_j - \sum_{j=1}^{n} \rho_{2j}W_j \right) \\ &= \sum_{j=1}^{\infty} (\rho_{1j}-\rho_{2j})W_j \\ &= 0 \end{aligned}$$

On the other hand, if $Y_1-Y_2 = 0$ almost surely, then by Theorem 2.4,

$$0 = \lim_{t\to\infty} \frac{P[|Y_1-Y_2| > t]}{P[|W_1| > t]} = \sum_{j=1}^{\infty} |\rho_{1j}-\rho_{2j}|^{\alpha}. \qquad \#$$

Therefore, we see that dispersion is not only a measure of tail thickness, but can also be used to define distance between two random variables in S. We remark that if, as in the example given in Section 2, $\sum_{j=1}^{\infty} |\rho_j|^{\alpha} < \infty$ is sufficient for existence, then the use of $\delta < \alpha$ is not required.

The obvious next question concerns the nature of convergence in this metric. We answer this partially.

Lemma 3.2 Convergence on S with the metric d implies convergence in ℓ_δ

Proof: Since $\delta<\alpha$ then $E|W_j|^{\delta} < \infty$. Furthermore, for any $Y = \sum_{j=1}^{\infty} \rho_j W_j$,

$$\begin{aligned} E|Y|^{\delta} &\le E\left[\sum_{j=1}^{\infty} |\rho_j W_j|^{\delta} \right] \\ &= E|W_1|^{\delta} \sum_{j=1}^{\infty} |\rho_j|^{\delta} \qquad \text{since } \delta \le 1 \qquad (3.1) \end{aligned}$$

Therefore, for $Y_n = \sum_{j=1}^{\infty} \rho_{nj} W_j$, $\mathrm{disp}(Y_n) \to 0$ implies $\sum_{j=1}^{\infty} |\rho_{nj}|^{\delta} \to 0$ which in turn implies $E|Y_n|^{\delta} \to 0$. #

We recognize that when $\alpha=2$, $\mathcal{S}$ is a subspace of a Hilbert space. In this case and when $\alpha>2$ so that variances are finite, it is usually most convenient to consider a Hilbert space setting. However, we are primarily interested in cases where $\alpha<2$. To consider projection operators in $\mathcal{S}$, let $X_1, \ldots, X_n \in \mathcal{S}$, then for any $Y \in \mathcal{S}$ define the projection operator $P_{\underline{X}}$ by

$$P_{\underline{X}} Y = \{\hat{Y} = \underline{a}'\underline{X} \text{ such that } \mathrm{disp}(Y-\hat{Y}) \text{ is minimum}\}.$$

<u>Theorem 3.3</u> Assume $\alpha>1$ and suppose $X_i = \sum_{j=1}^{\infty} \pi_{ij} W_j \in \mathcal{S}$, where for each $m \geq n$, $\Pi^{(m)} = [\pi_{ij}]_{i=1}^{n} {}_{j=1}^{m}$ is of full rank n. Suppose also that $Y = \sum_{j=1}^{\infty} \rho_j W_j \in \mathcal{S}$. Then $P_{\underline{X}} Y$ has a unique element. Furthermore, if $X_i^{(m)} = \sum_{j=1}^{m} \pi_{ij} W_j$, then

$$\hat{Y} = P_{\underline{X}} Y = \lim_{m\to\infty} P_{\underline{X}^{(m)}} Y.$$

<u>Proof</u>: We start by assuming $X_i = \sum_{j=1}^{m} \pi_{ij} W_j$, $Y = \sum_{j=1}^{m} \rho_j W_j$ where $n \leq m$ and m is finite. We wish to minimize

$$h(\underline{a}) = \mathrm{disp}(Y-\underline{a}'\underline{X})$$

$$= \sum_{j=1}^{m} |\rho_j - \underline{a}'\underline{\pi}_j|^{\alpha} \tag{3.2}$$

where $\underline{\pi}_j$ is the j^{th} column of $\Pi = [\pi_{ij}]_{i=1}^{n} {}_{j=1}^{m}$. We have assumed Π has full row rank n. Define

$$D_j = \{\underline{a}\varepsilon R^n \text{ such that } \underline{a}'\underline{\pi}_j = \rho_j\}$$

and $g_j(\underline{a}) = |\rho_j - \underline{a}'\underline{\pi}_j|^\alpha$.

For $\underline{a}\notin D_j$ (using $[x]^\gamma = \mathrm{sgn}(x)|x|^\gamma$),

$$\frac{\partial g_j(\underline{a})}{\partial \underline{a}} = \alpha\underline{\pi}_j[\underline{a}'\underline{\pi}_j - \rho_j]^{\alpha-1}$$

and
$$\frac{\partial^2 g_j(\underline{a})}{\partial \underline{a}\underline{a}'} = \alpha(\alpha-1)\underline{\pi}_j\underline{\pi}_j' \ |\underline{a}\underline{\pi}_j - \rho_j|^{\alpha-2}. \tag{3.3}$$

Since $\alpha>1$ and $\underline{\pi}_j\underline{\pi}_j'$ is nonnegative definite then g_j is convex off of D_j. In fact, g_j is minimized on D_j so that g_j is everywhere convex. We can actually go a step further and say that for $\underline{a}_1$, $\underline{a}_2\varepsilon R^n$, $\lambda\varepsilon(0,1)$,

$$\lambda g_j(\underline{a}_1)+(1-\lambda)g_j(\underline{a}_2) \geq g_j(\lambda\underline{a}_1+(1-\lambda)\underline{a}_2)$$

with equality iff $\underline{a}_1'\underline{\pi}_j = \underline{a}_2'\underline{\pi}_j$. That is, g_j is strictly convex except along lines orthogonal to $\underline{\pi}_j$. Equality cannot hold for every j, since π is full rank, so that $h = \sum_{j=1}^{m} g_j$ must be strictly convex. Furthermore, as $\max\limits_{1\leq j\leq n} |a_j|\to\infty$, $h(\underline{a})\to\infty$. Thus h must have a unique minimum.

The argument that h has a unique minimum holds even when the series are infinite, that is, when $Y = \sum_{j=1}^{\infty}\rho_j W_j$, $X_i = \sum_{j=1}^{\infty}\pi_{ij}W_j$ and

$$\begin{aligned} h(\underline{a}) &= \mathrm{disp}(Y-\underline{a}'\underline{X}) \\ &= \sum_{j=1}^{\infty}|\rho_j - \underline{a}'\underline{\pi}_j|^\alpha. \end{aligned}$$

Let $\underline{a}_0$ be the unique minimum of $h(\underline{a})$ and set (for $m \geq n$)

$h_m(\underline{a}) = \sum_{j=1}^{m} |\rho_j - \underline{a}'\pi_j|^{\alpha}$ with unique minimum $\underline{a}_m$. Suppose for some subsequence $\{\underline{a}_{m_k}\}$, $\max_{1 \leq j \leq n} |a_{m_k j}| \to \infty$. Then

$$\lim_{n\to\infty} h_{m_k}(\underline{a}_{m_k}) \geq \lim_{n\to\infty} h_n(\underline{a}_{m_k}) = \infty.$$

But for any m,

$$\begin{aligned} h_m(\underline{a}_m) &= \inf_{\underline{a}} h_m(\underline{a}) \\ &\leq \inf_{\underline{a}} h(\underline{a}) \\ &= h(\underline{a}_0). \end{aligned} \tag{3.4}$$

Thus, the sequence $\{\underline{a}_m\}$ must be compact and every subsequence must have a convergent subsequence. Suppose $\underline{a}_{m_k} \to \underline{a}_1$. Then from (3.4)

$$\begin{aligned} h_{m_k}(\underline{a}_{m_k}) &\leq h(\underline{a}_0) \\ &\leq h(\underline{a}_1). \end{aligned}$$

But since $h_m \uparrow h$ and the functions are all continuous, convergence is locally uniform, by an application of Dini's Theorem. This means

$$\lim_{k\to\infty} h_{m_k}(\underline{a}_{m_k}) = h(\underline{a}_1)$$

which implies $h(\underline{a}_1) = h(\underline{a}_0)$ and hence $\underline{a}_1 = \underline{a}_0$. Thus $\underline{a}_0 = \lim_{m\to\infty} \underline{a}_m$ and

$$P_{\underline{X}}Y = \underline{a}_0' \, \underline{X} = \lim_{m\to\infty} \underline{a}_m'X = \lim_{m\to\infty} P_{X^{(m)}}Y. \qquad \#$$

To actually calculate $\underline{a}_m$ is not easy unless either m=n ($\underline{a}_n = (\Pi^{(n)'})^{-1}\underline{\rho}$) or $\alpha=2$ ($\underline{a}_m = (\Pi^{(m)}\Pi^{(m)'})^{-1}\Pi^{(m)}\underline{\rho}$). An iterative procedure would be

$$\underline{a}_{m,1} = (\Pi^{(m)}\Pi^{(m)'})^{-1}\Pi^{(m)}\underline{\rho}$$

$$\underline{a}_{m,k+1} = \underline{a}_{m,k} - [(\Pi^{(m)}\Pi^{(m)'})^{-1}\Pi^{(m)}\underline{\ell}^{(m)}(\underline{a}_{m,k})]$$

where $\ell_j^{(m)}(\underline{a}) = [\underline{a}'\underline{\Pi}_j^{(m)} - \rho_j]^{\alpha-1}$ (using $[x]^{\ell} = \mathrm{sgn}(x)|x|^{\ell}$), $\underline{1 \le j \le m}$.

Even though the mapping $Y \to \hat{Y} = P_{\underline{X}}Y$ is unique it will not be a linear mapping (except when $\alpha=2$ or m=n). (See the example at the end of the section.)

<u>Theorem 3.4</u> Assume that $\underline{X}$ and Y are as in Theorem 3.3, except assume $\alpha \underline{<} 1$. To minimize disp($Y-\underline{a}'\underline{X}$), it suffices to consider $\underline{a}\varepsilon\bar{E}$, the closure of $E = \{\underline{a}\varepsilon R^n : \underline{a}'\underline{\pi}_j = \rho_j$ for at least n values of j$\}$.

<u>Proof</u>: As before, we seek to minimize

$$h^{(m)}(\underline{a}) = \sum_{j=1}^{m} |\rho_j - \underline{a}'\underline{\pi}_j|^{\alpha}, \quad m \underline{>} n. \tag{3.5}$$

Define again $D_j = \{\underline{a}\varepsilon R^n : \underline{a}'\underline{\pi}_j = \rho_j\}$ and $g_j(\underline{a}) = |\underline{a}'\underline{\pi}_j - \rho_j|^{\alpha}$. The matrix of second derivatives given in (3.3) indicate that at every $\underline{a} \notin D_j$, g_j is concave since $\alpha \underline{<} 1$. However, g_j is minimized on D_j. Since $h^{(m)} = \sum_{j=1}^{m} g_j$ is continuous everywhere, concave at all $\underline{a} \notin \bigcup_{j=1}^{m} D_j$ and infinite at infinity, then $h^{(m)}$ must therefore be minimized on $\bigcup_{j=1}^{m} D_j$. (This is not to say that points of minimum are exclusively in this set.)

Now consider the set

$$E_m = \{\underline{a}\varepsilon R^n : \underline{a}'\underline{\pi}_j = \rho_j \text{ for at least n values of } j\varepsilon\{1,2, \ldots,m\}\}.$$

Since $\Pi^{(m)} = [\pi_{ij}]_{i=1\ j=1}^{n\quad m}$ has rank n, then $D_j \cap E_m$ is non empty.

Suppose $\underline{a}_1 \varepsilon D_j$ and $\underline{a}_2 \varepsilon D_j \cap E_m$. From (3.5) we clearly have

$h^{(m)}(\underline{a}_1) \geq h^{(m)}(\underline{a}_2)$. Thus $h^{(m)}$ will be minimized on the set $\bigcup_{j=1}^{m} (D_j \cap E_m) = E_m$.

Suppose $\underline{a}_m \varepsilon E_m$ minimizes $h^{(m)} = \sum_{j=1}^{m} g_j$. To minimize $h = \sum_{j=1}^{\infty} g_j$ we consider the sequence $\{\underline{a}_m\}$. As in Theorem 3.3, this sequence must be compact, and hence there exists a subsequence $\underline{a}_{m_k} \to \underline{a}_0 \varepsilon \bar{E}$ where $E = \lim_{m\to\infty} E_m$. That $\underline{a}_0$ will minimize h is also true, and this is argued as in the previous theorem. #

The point of minimum $\underline{a}_m$ for $h^{(m)}$ will not necessarily be unique, except when m=n. In that case, $\underline{a}_m = (\Pi^{(m)'})^{-1}\underline{\rho}$ and the mapping $Y \to P_{\underline{X}} Y$ is linear. When m=n+1, however, such a linear mapping can still be defined, even when there is not a unique minimum.

<u>Theorem 3.5</u> Suppose $Y = \sum_{j=1}^{\infty} \rho_j W_j \varepsilon S$, $\alpha \leq 1$ and suppose, for $1 \leq i \leq n$, $X_i = \sum_{j=1}^{n+1} \pi_{ij} W_j$ and $\Pi = [\pi_{ij}]$ has rank n. Then there exists a linear mapping $Y \to \hat{Y}$ into span $\{X_1, \ldots, X_n\}$ which minimizes $\mathrm{disp}(Y-\hat{Y})$.

Proof: Let $Z = \sum_{j=n+2}^{\infty} \rho_j W_j$ so that $Y = Z+\underline{\rho}'\underline{W}$, where $\underline{\rho} = (\rho_1, \ldots, \rho_{n+1})$ and $\underline{W} = (W_1,\ldots, W_{n+1})$.

We wish to minimize

$$\begin{aligned} h(\underline{a}) &= \mathrm{disp}(Y-\underline{a}'\underline{X}) \\ &= \mathrm{disp}(Z) + \sum_{j=1}^{n+1} |\rho_j-\underline{a}'\underline{\pi}_j|^{\alpha}. \end{aligned}$$

Since $\alpha\leq 1$, then according to Theorem 3.4 a solution is given by $\underline{a}$ satisfying $\underline{a}'\underline{\pi}_j = \rho_j$ for at least n values of $j \in \{1, 2, \ldots, n+1\}$. If $\underline{\rho} = \Pi'\underline{a}_0$ for some $\underline{a}_0 \in R^n$, then $Y = Z+\underline{a}_0'\underline{X}$ and Lemma 2.1 applies. In this case, $\hat{Y} = \underline{a}_0'\underline{X}$ is the unique solution and the mapping is linear.

On the other hand, if $\underline{\rho}$ is not in the row space of Π, then it suffices to consider $\underline{a}$ such that $\underline{a}'\underline{\pi}_j = \rho_j$ for exactly n values of j. Suppose k is the one value for which $\underline{a}'\pi_k \neq \rho_k$. Define Π_{-k} and $\underline{\rho}_{-k}$ to be Π and $\underline{\rho}$, respectively, with k^{th} column ($\underline{\pi}_k$) and k^{th} element removed. Then $\underline{a} = (\Pi'_{-k})^{-1}\underline{\rho}_{-k}$ and

$$\begin{aligned} \min_{\underline{a}} h(\underline{a}) &= \min_{\underline{a}} \sum_{j=1}^{n+1} |\rho_j-\underline{a}'\underline{\pi}_j|^{\alpha}+\mathrm{disp}(Z) \\ &= \min_{1\leq k\leq n+1} |\rho_k-\underline{\rho}'_{-k}(\Pi_{-k})^{-1}\underline{\pi}_k|^{\alpha}+\mathrm{disp}(Z) \qquad (3.6) \end{aligned}$$

By inverting $Q = [\Pi'\underline{\rho}]$ we can find the $(n+1,k)$ element of Q^{-1}

$$[\rho_k-\underline{\rho}'_{-k}(\Pi_{-k})^{-1}\underline{\pi}_k]^{-1} = (-1)^{k+n+1}\det(\Pi_{-k})[\det(Q)]^{-1}. \qquad (3.7)$$

Note that this factors into a part depending only on $\underline{\rho}$ and a part depending only on k.

Define $j_0 = \max\{k \leq n+1 : |\det(\Pi_{-k})^{-1}|$ is minimum . Then from (3.6) and (3.7) we have

$$\min_{\underline{a}} h(\underline{a}) = \left|\frac{\det(Q)}{\det(\Pi_{-j_0})}\right|^{\alpha} + \mathrm{disp}(Z)$$

$$= |\rho_{j_0} - \underline{\rho}'_{-j_0}(\Pi_{-j_0})^{-1}\underline{\Pi}_{j_0}|^{\alpha} + \mathrm{disp}(Z).$$

And a point of minimum for h is $\underline{a}_0 = (\Pi'_{-j_0})^{-1}\underline{\rho}_{-j_0}$. If we define P to be the matrix $(\Pi'_{-j_0})^{-1}$ where a column of zeroes is squeezed in to make a new j_0 th column, then $\underline{a}_0 = P\underline{\rho}$.

By this definition, $\hat{Y} = (P\underline{\rho})'\underline{X}$ defines a linear mapping on S and $\hat{Y} \varepsilon P_{\underline{X}} Y$, so that it is a minimum dispersion predictor for Y. #

The following example illustrates that $P_{\underline{X}}$ is not necessarily linear. With $X = W_1+2W_2+W_3$, then $P_X(W_1+W_2) \neq P_X(W_1)+P_X(W_2)$ for $\alpha = 1/2, 1, 3/2$.

α	1/2	1	3/2
$P_X(W_1)$	0	0	$\frac{2-\sqrt{2}}{4}$ X
$P_X(W_2)$	0	0	1/3 X
$P_X(W_1+W_2)$	1/2 X	1/2 X	1/2 X

4. ARMA Processes With Regularly Varying Tails

We now look specifically at autoregressive - moving average processes driven by a sequence $\{W_j\}_{j=-\infty}^{\infty}$, independent and of identical, regularly varying distribution F*. As before, let

$\bar{F}(t) = P[|W_j|>t]\varepsilon \mathcal{R}V_{-\alpha}$, $\alpha>0$.

<u>Theorem 4.1</u> There exists a stationary sequence $\{X_n\}_{n=-\infty}^{\infty}$ satisfying

$$X_n-\phi_1X_{n-1}-\dots-\phi pX_{n-p} = W_n+\theta_1W_{n-1}+\dots+\theta_qW_{n-q} \tag{4.1}$$

for all n, if $\Phi(z) = 1-\phi_1z-\dots-\phi_pz^p \neq 0$ for all complex z such that $|z| = 1$. Furthermore, suppose the sequence $\{\tau_n\}_{n=-\infty}^{\infty}$ and generating function $T(z) = \sum_{n=-\infty}^{\infty} \tau_nz^n$ satisfy $T(z)\dfrac{1+\theta_1z+\dots+\theta_qz^q}{1\underset{-\phi_1}{\cancel{+\theta_1}}z-\dots\underset{-\phi_p}{\cancel{-\theta_p}}z^p} \neq 0$ for $1-\varepsilon<|z|<1+\varepsilon$, for some $\varepsilon>0$. Then the infinite series $\sum_{n=-\infty}^{\infty} \tau_nX_n$? is almost surely absolutely convergent and has regularly varying tails equivalent to $\bar{F}$.

<u>Proof</u>. Since $\Phi(z) \neq 0$ for $|z| = 1$, there exists $\zeta\varepsilon(0,1)$ and a sequence $\{\sigma_j\}$ such that $\Sigma(z) = \sum_{j=-\infty}^{\infty} \sigma_jz^j = \dfrac{1}{\Phi(z)}$ for complex z such that

$\zeta<|z|<\zeta^{-1}$. In particular, $|\sigma_j|\leq k\zeta^{-|j|}$ for all j and some k>0 and

$$\sigma_j-\phi_1\sigma_{j-1}-\dots-\phi p\sigma_{j-p} = \begin{cases} 0 & \text{if } j\neq0 \\ 1 & \text{if } j=0 \end{cases}. \tag{4.2}$$

By Theorem 2.1, $\sum_{j=-\infty}^{\infty} \sigma_j W_{n-j}$ is almost surely absolutely convergent for all n.

Let $\pi_j = \sigma_1+\theta_1\sigma_{j-1}+\ldots+\theta_q\sigma_{j-q}$ and $X_n = \sum_{j=-\infty}^{\infty} \pi_j W_{n-j}$. Then X_n exists as well and $\{X_n\}$ is stationary. Furthermore

$$X_n = \sum_{j=-\infty}^{\infty} \sigma_j(W_{n-j}+\theta_1 W_{n-1-j}+\ldots+\theta_q W_{n-q-j}).$$

Using (4.2), we show that $\{X_n\}$ satisfies (4.1).

$$\begin{aligned} X_n-\phi_1 X_{n-1}-\ldots-\phi_p X_{n-p} &= \sum_{j=-\infty}^{\infty} (\sigma_j-\phi_1\sigma_{j-1}-\ldots-\phi_p\sigma_{j-p})(W_{n-j} \\ &\qquad +\theta_1 W_{n+1-j}+\ldots+\theta_q W_{n-q-j}) \\ &= W_n+\theta_1 W_{n-1}+\ldots+\theta_q W_{n-q}. \end{aligned}$$

Now suppose T(z) is as in the statement of the theorem. To show that $\sum_{n=-\infty}^{\infty} \tau_n X_n$ is almost surely absolutely convergent, we need to show that $\sum_{j=-\infty}^{\infty} \left| \sum_{n=-\infty}^{\infty} \tau_n \pi_{n-j} \right|^{\delta} <\infty$ for some $\delta<\alpha$, $\delta\leq 1$. It suffices to show that $\left| \sum_{n=-\infty}^{\infty} \tau_n \pi_{n-j} \right|$ decreases at least geometrically as $|j|\to\infty$. But this is indeed the case, since

$$\sum_{j=-\infty}^{\infty} \left(\sum_{n=-\infty}^{\infty} \tau_n \pi_{n-j} \right) z^j = T(z)\, \frac{1+\theta_1 z+\ldots+\theta_q z^q}{1-\phi_1 z-\ldots-\phi_p z^p} \neq 0$$

for any complex z in an annulus containing the unit circle. #

This theorem, of course, is proven almost identically to the analogous theorem for finite variance processes. The same is true of the next theorem. We first provide a couple of definitions. A stationary sequence $\{X_n\}$ satisfying (4.1) is said to be causal if for some sequence $\{\pi_j\}_{j=0}^{\infty}$ and some $\delta<\alpha$, $\delta\leq 1$, $X_n = \sum_{j=0}^{\infty} \pi_j W_{n-j}$ almost surely and $\sum_{j=0}^{\infty} |\pi_j|^{\delta}<\infty$. $\{X_n\}$ is said to be invertible if for some sequence $\{\psi_j\}_{j=0}^{\infty}$ and some $\delta<\alpha$, $\delta\leq 1$, $W_n = \sum_{j=0}^{\infty} \psi_j X_{n-j}$ almost surely and $\sum_{j=0}^{\infty} |\psi_j|^{\delta}<\infty$.

<u>Theorem 4.2</u> Suppose $\{X_n\}$ is a stationary sequence satisfying (4.1) where $\Phi(z) = 1-\phi_1 z-\ldots-\phi_p z^p$ and $\Theta(z) = 1+\theta_1 z+\ldots+\theta_q z^q$ do not have common roots. Then

i) $\{X_n\}$ is causal if and only if $\Phi(z) \neq 0$ for every complex z such that $|z|\leq 1$.

ii) $\{X_n\}$ is invertible if and only if $\Theta(z) \neq 0$ for every complex z such that $|z|\leq 1$.

<u>Proof</u>: Let $\theta_j=0$ for $j>q$ and $\phi_j=0$ for $j>p$, and $\theta_0=\phi_0=1$.

i) Suppose $\{X_n\}$ is causal, that is, $X_n = \sum_{j=1}^{\infty} \pi_j W_{n-j}$ and $\sum_{j=1}^{\infty} |\pi_j|^{\delta}<\infty$ for some $\delta<\alpha$, $\delta\leq 1$. Define $\Pi(z) = \sum_{j=0}^{\infty} \pi_j z^j$ and $\{\sigma_j\}$ so that $\Sigma(z) = \Phi(z)\Pi(z) = \sum_{j=0}^{\infty} \sigma_j z^j$. Clearly $\sum_{j=0}^{\infty} |\sigma_j|^{\delta}<\infty$ and $\sum_{j=0}^{\infty} |\sigma_j-\theta_j|^{\delta}<\infty$.

Therefore

$$\sum_{j=0}^{\infty} \theta_j W_{n_j} = X_n-\phi_1 X_{n-1}-\ldots-\phi_p X_{n-p} = \sum_{j=0}^{\infty} \sigma_j W_{n-j}$$

and hence $0 = \text{disp} \sum_{j=0}^{\infty} (\sigma_j-\theta_j)W_{n-j} = \sum_{j=0}^{\infty} |\sigma_j-\theta_j|^{\delta}$. This implies $\Sigma(z)=\Theta(z)$. Since $\Pi(z)$ has no poles when $|z|\leq 1$ and $\Theta(z)$ has no roots in common with $\Phi(z)$, then $\Phi(z) = \frac{\Theta(z)}{\Pi(z)}$ has no roots in $|z|\leq 1$.

Now suppose $\Phi(z)\neq 0$ for $|z|\leq 1$. Then there exists $\zeta<1$ such that the roots of $\Phi(z)$ are in $|z|>\zeta^{-1}$. Let $\Pi(z) = \sum_{j=0}^{\infty} \Pi_j z^j = \frac{\Theta(z)}{\Phi(z)}$. For some $k>0$, $|\pi_j|\leq k\zeta^j$ for all j and $\sum_{j=0}^{\infty} |\pi_j|^\delta<\infty$ for any $\delta>0$. Therefore, the series

$Y_n = \sum_{j=0}^{\infty} \pi_j W_{n-j}$ is almost surely absolutely summable for every n. Since $\pi_j-\phi_1\pi_{j-1}-\ldots-\phi_p\pi_{j-p}=\theta_j$ for all $j>0$, it follows that $\{Y_n\}$ satisfies (4.1). Defining $\Sigma(z) = \sum_{j=0}^{\infty} \sigma_j z^j = \frac{1}{\Phi(z)}$, then $\sum_{=0}^{\infty} |\sigma_j|^\delta<\infty$ and

$\pi_j = \sigma_j+\theta_1\sigma_{j-1}+\ldots+\theta_q\sigma_{j-q}$. Thus,

$$
\begin{aligned}
X_n &= \sum_{j=0}^{\infty} \sigma_j (W_{n-j}+\theta_1 W_{n-1-j}+\ldots+\theta_p W_{n-p-j}) \\
&= \sum_{j=0}^{\infty} \pi_j W_{n-j} \\
&= Y_n .
\end{aligned}
$$

ii) Suppose $\{X_n\}$ is invertible, so that $W_n = \sum_{j=0}^{\infty} \psi_j X_{n-j}$ where $\sum_{j=0}^{\infty} |\psi_j|^\delta<\infty$ for some $\delta<\alpha$, $\delta\leq 1$. Define $\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$ and $\Sigma(z) = \sum_{j=0} \sigma_j z^j = \Theta(z)\Psi(z)$. Then $\sum_{j=0}^{\infty} |\sigma_j|^\delta<\infty$. Using absolute convergence to rearrange series,

$$
\begin{aligned}
W_n-\phi_1 W_{n-1}-\ldots-\phi_p W_{n-p} &= \sum_{j=0}^{\infty} \psi_j (X_{n-j}-\phi_1 X_{n-1-j}-\ldots-\phi_p X_{n-p-j}) \\
&\sum_{j=0}^{\infty} \psi_j (W_{n-j}+\theta_1 W_{n-1-j}+\ldots+\theta_q W_{n-q-j}) \\
&\sum_{j=0}^{\infty} \sigma_j W_{n-j} .
\end{aligned}
$$

This implies $\Sigma(z) = \Phi(z)$, because $0 = \text{disp} \sum_{j=0}^{\infty} (\sigma_j - \phi_j) W_{n-j}$ $= \sum_{j=0}^{\infty} |\sigma_j - \phi_j|^{\delta}$. Since $\Psi(z)$ has no poles in $|z| \leq 1$ and since $\Phi(z)$ and $\Theta(z)$ have no common roots, then $\Theta(z) = \frac{\Phi(z)}{\Psi(z)}$ has no roots in $|z| \leq 1$.

Finally, suppose $\Theta(z) \neq 0$ in $|z| \leq 1$, and define $\Sigma(z) = \frac{1}{\Theta(z)}$. The coefficients $\{\sigma_j\}$ of $\Sigma(z)$ will decrease at least as fast as a geometric sequence, so $\sum_{j=0}^{\infty} |\sigma_j|^{\delta} < \infty$ for any $\delta > 0$. Then $\delta_j + \theta_1 \sigma_{j-1} + \ldots + \theta_q \sigma_{j-q} = 0$, $j \neq 0$. So, for any n,

$$W_n = \sum_{j=0}^{\infty} \sigma_j (W_{n-j} + \theta_1 W_{n-1-j} + \ldots + \theta_q W_{n-q-j})$$

$$= \sum_{j=0}^{\infty} \sigma_j (X_{n-j} - \phi_1 X_{n-1-j} - \ldots - \phi_p X_{n-p-j})$$

$$= \sum_{j=0}^{\infty} \psi_j X_{n-j},$$

where $\psi_j = \sigma_j - \phi_1 \sigma_{j-1} - \ldots - \phi_p \sigma_{j-p}$. #

If the sequence $\{X_n\}$ satisfies (4.1) and is stationary, causal, then the expression $X_n = \sum_{j=0}^{\infty} \pi_j W_{n-j}$ gives us the dispersion of X_n.

$$\text{disp}(X_n) = \sum_{j=0}^{\infty} |\pi_j|^{\alpha}.$$

And in particular, if $\hat{X}_{n+1} = a_1 X_n + \ldots + a_n X_1$ is the predicted value of $X_{n+1} = \sum_{j=-\infty}^{\infty} \pi_j W_{n+1-j}$, then

$$\operatorname{disp}(\hat{X}_{n+1}-X_{n+1}) = 1 + \sum_{j=0}^{\infty} |\pi_{n+1-j}-(a_1\pi_{n-j}+\ldots+a_n\pi_{1-j})|^{\alpha} .$$

If $a_1, \ldots, a_n$ are chosen so that this error dispersion is minimized, then $\hat{X}_{n+1}$ will be an element in the projection of X_{n+1} onto span $\{X_1, \ldots, X_n\}$, as described in Section 2. Part II discusses in detail the prediction problem for ARMA processes with regularly varying tails.

PART II: LINEAR PREDICTION OF ARMA PROCESSES WITH INFINITE VARIANCES

Summary To predict unobserved values of a linear process with infinite variance, we introduce a linear predictor which minimizes the chance of large prediction errors. The procedure corresponds to minimizing, in a linear space setting, an ℓ_α $(0<\alpha<2)$ distance between predicted and actual values, and is the natural procedure when the process is driven by symmetric stable noise. We derive explicitly the best linear predictor of X_{n+1} in terms of $X_1, \ldots, X_n$ for the process ARMA (1,1) and for the process AR(p). For higher order processes, general analytic expressions are cumbersome, but we indicate how predictors can be determined numerically. Numerical comparisons with the least squares predictor accompany the report.

1. Introduction

To define the problem we express the ARMA (p, q) process $\{X_n\}_{-\infty}^{\infty}$ as the stationary solution of

$$X_n-\phi_1X_{n-1}-\ldots-\phi_pX_{n-p} = W_n+\theta_1W_{n-1}+\ldots+\theta_qW_{n-q} \tag{1.1}$$

where $\{W_n\}$ are independent and identically distributed and

$$(1-\phi_1z-\ldots-\phi_pz^p)(1+\theta_1z+\ldots+\theta_qz^q) \neq 0 \text{ for complex } |z|\leq 1. \tag{1.2}$$

For purposes of this investigation, the residuals $\{W_n\}_{n=-\infty}^{\infty}$ will have distribution which satisfies for all $x>0$

$$\lim_{t\to\infty} \frac{P[|W_n|>xt]}{P[|W_n|>t]} = x^{-\alpha}, \quad \alpha>0. \tag{1.3}$$

Such distributions are said to have regularly varying tails and the parameter α is called the tail index. When $\alpha<2$, the variance does not exist. We also assume throughout the main body of this paper that W_n has distribution symmetic about zero. This provides a precise notion of location of a distribution.

Part I of the thesis discusses infinite series of regularly varying tailed random variables and demonstrates that a stationary solution to (1.1) does exist when (1.2) holds. This solution satisfies

$$X_n = \sum_{j=-\infty}^{n} \pi_{n-j} W_j$$

where the $\{\pi_j\}_{j=0}^{\infty}$ are determined (as in the usual manner for finite variance ARMA processes) from the coefficients $\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q$ in (1.1).

We desire to predict the future values of the process, X_{n+1}, $X_{n+2},\ldots$ based on the observed data $X_1, X_2, \ldots, X_n$. Since the process is linear, we consider only linear predictors. The predictor for a random variable Y will be denoted $\hat{Y} = \underline{a}'\underline{X}_n$ where $\underline{a}' = (a_1,\ldots, a_n)$ and $\underline{X}_n' = (X_n,X_{n-1},\ldots,X_1)$.

Traditionally, linear prediction for a stationary ARMA (p,q) process (satisfying (1.1) and (1.2)) has relied on the least squares procedure (see e.g. Fuller (1976) and Box and Jenkins (1976)). When the process has finite variance, the least squares predictor minimizes the mean squared prediction error and when the process is Gaussian the procedure has maximum concentration. If the variance is infinite, however, least squares may be ineffective and perhaps even inappropriate. A different measure of prediction error becomes necessary, although we may be forced to relinquish the elegant Hilbert space results of the least squares predictor. Other possibilities spring to mind, for example minimum mean absolute deviation prediction or a pseudo-spectral approach as that studied by Cambanis and Soltani (1982). Most such measures are extremely unwieldy in practice, however, and require precise knowledge of the noise distribuiton. One would prefer a predictor which can be calculated with only minimal knowledge of the noise distribution. In addition, infinite variance processes can have quite extraordinary, outlying values and this suggests minimizing the probability that a large error occurs in prediction.

Fortunately just such a predictor exists for symmetric stable processes. This predictor is obtained by use of the "natural" criterion - minimizing the spread of the prediction error distribution. Stuck (1978) has quite successfully utilized this criterion, which he calls minimum dispersion, for Kalman filtering problems with symmetric stable processes. Recognizing that for any sequence $\{W_j\}$ of independent and identically symmetric stable (α) random variables, $Y = \sum_{j=1}^{\infty} \rho_j W_j$ is also symmetric stable (α) with relative scale $\left(\sum_{j=1}^{\infty} |\rho_j|^{\alpha}\right)^{1/\alpha}$, i.e., $Y \overset{d}{=} \left(\sum_{j=-\infty}^{\infty} |\rho_j|^{\alpha} W_j\right)^{1/\alpha}$, Stuck defined the dispersion of Y as

$$\mathrm{disp}(Y) = \sum_{j=-\infty}^{\infty} |\rho_j|^{\alpha}. \tag{1.4}$$

This extends the usual notion of dispersion (variance) for Gaussian variables. Regardless of the distribution of W_n, if it satisfies (1.3) then (1.4) will define the dispersion of $Y = \sum_{j=-\infty}^{\infty} \rho_j W_j$ for us.

Writing the ARMA process $\{X_n\}$ in its moving average form $X_n = \sum_{j=-\infty}^{n} \pi_{n-j} W_j$ and letting $\hat{Y} = a_1 X_n + \ldots + a_n X_1$ we define the minimum error dispersion linear predictor of $Y = \sum_{j=-\infty}^{\infty} \rho_j W_j$ to be that $\hat{Y}$ which minimizes (over all choices of $a_1, \ldots, a_n$)

$$\mathrm{disp}(\hat{Y}-Y) = \sum_{j=-\infty}^{\infty} |\rho_j - (a_1 \pi_{n-j} + \ldots + a_n \pi_{1-j})|^{\alpha}. \tag{1.5}$$

In the special case where $Y=X_{n+k}$ we minimize

$$\mathrm{disp}(\hat{X}_{n+k}-X_{n+k}) = 1+\sum_{j=-\infty}^{n} |\pi_{n+k-j}-(a_1\pi_{n-j}+\ldots+a_n\pi_{1-j})|^{\alpha}. \tag{1.6}$$

For a linear process driven by symmetric stable noise, the prediction error for any linear predictor also has symmetric stable distribution. The minimum dispersion prediction error has the distribution with the smallest scale and hence is optimal. The procedure is easily extended to more general linear processes, since it requires only knowledge of the coefficients of the process and of the tail index α of the noise distribution. Furthermore, by use of the next lemma we can relate dispersion to the probability of large values. The corollary to this is that among linear predictors, the minimum dispersion predictor is optimal in the sense that it minimizes the probability of large prediction errors.

Lemma 1.1 Suppose $\{W_j\}$ are independent and identically distributed and satisfy (1.3) and suppose $Y = \sum_{j=-\infty}^{\infty} \rho_j W_j$ where $\sum_{j=-\infty}^{\infty} |\rho_j|^{\delta}<\infty$ for some $\delta<\alpha$, $\delta\leq 1$. Then

$$\lim_{t\to\infty} \frac{P[|Y|>t]}{P[|W|>t]} = \mathrm{disp}(Y) = \sum_{j=-\infty}^{\infty} |\rho_j|^{\alpha}.$$

Proof: See Part I, Theorem 2.4. #

Since the coefficients $\{\pi_j\}$ for an ARMA process are geometrically decreasing in magnitude, this lemma indicates that $\mathrm{disp}(\hat{X}_{n+k}-X_{n+k})$ is roughly proportional to the probability of a large prediction error.

We recognize disp$(\hat{Y}-Y)$ in (1.4) as being related to an ℓ_α type of distance between Y and $\hat{Y}$ on the linear space generated by $\{W_n\}$. In case $\alpha=2$, it is the usual Euclidean squared distance. In that case, $\hat{Y} = P_nY$ where P_n is a linear projection mapping Y form a Hilbert space into span $\{X_n,\ldots, X_1\}$, the space generated by linear combinations of $X_n, \ldots, X_1$. With $\alpha<2$, P_n is still an operator but not necessarily a linear operator $(P_n(Y_1+Y_2) \neq P_n(Y_1) + P_n(Y_2))$ and not even necessarily unique if $\alpha\leq 1$. (See Part I, Section 3 for a discussion of some of the linear space properties of P_n.)

We shall see in Sections 2 and 3 that minimum dispersion linear predictors can be found quite explicitly for autoregressive processes and for the mixed ARMA (1,1) process. In both cases, the prediction operator, P_n is unique and linear on span $\{X_1,X_2, \ldots\}$. For higher order moving average and mixed processes, however, one cannot always give a single general expression which is acceptable for all values of the parameters. For particular values, determination of the predictor is straightforward. Section 4 discusses the higher order processes and in Section 5 we compare numerically the minimum dispersion predictor with the least squares predictor.

2. Minimum Dispersion Prediction for Autoregressive Processes

For a purely autoregressive process, the difference equation (1.1) reduces to

$$X_n = \phi_1 X_{n-1}+\ldots+\phi_p X_{n-p}+W_n \tag{2.1}$$

where $\{W_n\}$ are independent, identically and symmetrically distributed and satisfy (1.3).

Assumption (1.2) guarantees that $X_n = \sum_{j=0}^{\infty} \pi_j W_{n-j}$ where the π_j are uniquely determined by

$$\sum_{j=0}^{\infty} \pi_j z^j = \left(\sum_{j=0}^{n} \phi_j z^j \right)^{-1}, \quad |z| \leq 1. \tag{2.2}$$

Furthermore (see Part, I, Section 4) the sequence $\{X_n\}$ is stationary and has symmetric marginal distributions with regularly varying tails with index α and dispersion

$$\operatorname{disp}(X_n) = \sum_{j=0}^{\infty} |\pi_j|^{\alpha} \tag{2.3}$$

Similarly, for any $\underline{a} = (a_1,a_2,\ldots,a_n)$, then $X_{n+k}-\underline{a}'\underline{X}_n = X_{n+k}-(a_1X_n+\ldots+a_nX_1)$ also has tail index α. We seek to find an $\underline{a}$ such that $\operatorname{disp}(X_{n+k}-\underline{a}'\underline{X}_n)$ is minimized. First we establish a useful lemma.

<u>Lemma 2.1</u> Let S_* be the class of random variables of the form $Y = Z+\underline{m}'\underline{X}_n$ for some $\underline{m}\in R^n$ and $Z = \sum_{j=n+1}^{\infty} \rho_j W_j$ such that Z exists. Then for each $Y \in S_*$, the set $P_n Y = \{\underline{a}'X_n : \operatorname{disp}(Y-\underline{a}'X_n) \text{ is minimum}\}$ consists of

exactly one variable. For $Y = Z+\underline{a}'_0\underline{X}_n$, this unique variable is $\hat{Y} = \underline{a}_0'\underline{X}_n$. Furthermore, the mapping $Y\to\hat{Y}$ is linear on S_*.

Proof: The dispersion of the prediction error $Y-\underline{a}'\underline{X}_n$ is

$$\begin{aligned}\mathrm{disp}(Y-\underline{a}'X_n) &= \mathrm{disp}(Z+(\underline{a}_0-\underline{a})'\underline{X}_n)\\ &= \sum_{j=n+1}^{\infty}|\rho_j|^{\alpha} + \sum_{j=-\infty}^{n}|(a_1-a_{01})\pi_{n-j}+\ldots+(a_n-a_{0n})\pi_{1-j}|^{\alpha}\\ &\geq \sum_{j=n+1}^{\infty}|\rho_j|^{\alpha}\end{aligned}$$

with equality if and only if $\underline{a}=\underline{a}_0$. Thus the minimum dispersion linear predictor is $\hat{Y} = \underline{a}_0'\underline{X}_n$ as asserted. The linearity of the mapping $Y = Z+\underline{a}_0'\underline{X}_n \to \hat{Y} = \underline{a}_0'\underline{X}_n$ is apparent from the form of Y. #

Corollary 2.2 For the process (2.1), provided $n\geq p$, there exists a unique minimum dispersion linear predictor $\hat{X}_{n+k}$ for X_{n+k} $(k\geq 1)$ in terms of $X_1,\ldots, X_n$. This predictor satisfies the recursive relationship

$$\hat{X}_{n+k} = \phi_1\hat{X}_{n+k-1}+\ldots+\phi_p\hat{X}_{n+k-p} \tag{2.4}$$

with initial conditions $\hat{X}_j = X_j$ for $1\leq j\leq n$.

Proof. We observe that each of $X_1,X_2,\ldots,X_n,X_{n+1}\ldots$ belong to the class S_* defined in Lemma 2.1. Since

$$X_{n+k} = W_{n+k}+\phi_1X_{n+k-1}+\ldots+\phi_pX_{n+k-p}$$

and $\hat{W}_{n+k} = 0$ by Lemma 2.1, then the linearity of the prediction mapping gives the relationship (2.4). #

Remarks

1. The minimum dispersion predictor is exactly the same as the least squares predictor $\tilde{X}_{n+k}$ for an autoregressive process. This is not the case for more general ARMA processes.

2. The residuals $W_{n+1}, W_{n+2}, \ldots$ are predicted with zeroes, and for $p<j\leq n$, then

$$\hat{W}_j = X_j-\phi_1 X_{j-1}-\ldots-\phi_p X_{j-p},$$

but the linearity principle does not apply to $W_1,\ldots,W_p$. In fact, if $\alpha\leq 1$, the set $P_n W_j = \{\underline{a}'\underline{X}_n : \text{disp}(W_j-\underline{a}'\underline{X}_n) \text{ is minimum}\}$ may not consist of only one element for $j\leq p$.

3.0 Prediction of the ARMA(1,1) Process

In this section we are concerned with the stationary process $\{X_n\}$ defined by

$$X_n - \phi X_{n-1} = W_n + \theta W_{n-1} \tag{3.1}$$

where $|\phi|<1$ and $|\theta|<1$ and $\{W_n\}$ are iid, satisfying (1.3). We find it necessary to distinguish between the cases $\alpha\leq 1$ and $\alpha>1$. For both cases, however, we shall need the following lemma.

Lemma 3.1 If $a>0$ and $\alpha>0$, then $h(x) = a|x|^\alpha + |x-b|^\alpha$ has its minimum value at x_m, where

$$x_m = \begin{cases} b & \text{if } \alpha\leq 1,\ a\leq 1 \\ 0 & \text{if } \alpha\leq 1,\ a>1 \\ \dfrac{b}{1+a^{1/\alpha-1}} & \text{if } \alpha>1 \end{cases}$$

and x_m is unique if $a\neq 1$ or $\alpha>1$.

The minimum value of h is

$$h(x_m) = \begin{cases} |b|^\alpha \min(1,a) & \text{if } \alpha\leq 1 \\ a|b|^\alpha (1+a^{1/\alpha-1})^{1-\alpha} & \text{if } \alpha>1. \end{cases}$$

Proof: Define the function $[x]^\gamma = \text{sgn}(x)|x|^\gamma$. Suppose $b>0$. Then for $x \neq 0$, $x \neq b$

$$h'(x) = \alpha(a[x]^{\alpha-1}+[x-b]^{\alpha-1})$$
$$h''(x) = \alpha(\alpha-1)(a|x|^{\alpha-2}+|x-b|^{\alpha-2}).$$

So for $x<0$, $h'(x)<0$ and for $x>b$, $h'(x)>0$. Thus h is minimized in $[0,b]$.

If $\alpha \leq 1$, then $h''(x) \leq 0$, so the minimum must be either at 0 or at b. It is easy to see that $h(b) \leq h(0)$ if and only if $a \leq 1$.

If $\alpha>1$, then h' is continuous on $[0,b]$ and h'' is positive. Thus $h'(x_m) = 0$ gives us the point of minimum. On $[0,b]$, $h'(x) = \alpha(ax^{\alpha-1} -(b-x)^{\alpha-1})$, so that $x_m = b(1+a^{1/\alpha-1})^{-1}$. Also

$$h(x_m) = a\left(\frac{b}{1+a^{1/\alpha-1}}\right)^\alpha + \left(\frac{ba^{1/\alpha-1}}{1+a^{1/\alpha-1}}\right)^\alpha$$
$$= \frac{ab}{(1+a^{1/\alpha-1})^{\alpha-1}} .$$

The proof is similar if $b<0$. #

We make use of this lemma first to deal with the case when $\alpha \leq 1$.

Theorem 3.2 For the ARMA(1,1) process (3.1) with $\alpha \leq 1$, a minimum dispersion linear predictor for X_{n+k} $(k \geq 1)$ based on $\underline{X}_n = (X_n,\ldots,X_1)$, is $\hat{X}_{n+k} = \underline{a}'\underline{X}_n$ where

$$a_j = (\phi+\theta)(-\theta)^{j-1}\phi^{k-1} \qquad , \ 1\le j\le n-1$$

$$a_n = \begin{cases} (\phi+\theta)(-\theta)^{n-1}\phi^{k-1} & \text{if } |\phi+\theta|^{\alpha} \le 1-|\phi|^{\alpha} \\ \phi^{k}(-\theta)^{n-1} & \text{if } |\phi+\theta|^{\alpha} \ge 1-|\phi|^{\alpha} \end{cases} \tag{3.2}$$

If $|\phi+\theta|^{\alpha} \neq 1-|\phi|^{\alpha}$, the predictor is unique.

The minimum value of the error dispersion is

$$\text{disp}(X_{n+k}-\underline{a}'\underline{X}_n) = 1+|\phi+\theta|^{\alpha} \ \frac{1-|\phi|^{\alpha(k-1)}}{1-|\phi|^{\alpha}}$$

$$+ \ |\phi|^{\alpha(k-1)}|\theta|^{n\alpha}\min\left(1, \ \frac{|\phi+\theta|^{\alpha}}{1-|\phi|^{\alpha}}\right).$$

<u>Proof</u>: Since $|\phi|<1$, we have

$$X_j = W_j+(\phi+\theta)\sum_{k=1}^{\infty}\phi^{k-1}W_{j-k} \qquad \text{for all } j. \tag{3.3}$$

If $\underline{m}\varepsilon R^n$ and if we define $m_0=-\phi^{k-1}$, then from (3.3) we can write for $k\ge 1$,

$$\underline{m}'\underline{X}_n-X_{n+k} = -W_{n+k}-\sum_{j=1}^{k-1}(\phi+\theta)\phi^{k-j-1}W_{n+j}$$

$$+ \sum_{j=1}^{n}\left[m_j+(\phi+\theta)\sum_{i=0}^{j-1} m_j\phi^{j-i-1}\right]W_{n+1-j}$$

$$+ (\phi+\theta)\left(\sum_{i=0}^{n} m_i\phi^{n-i}\right)\sum_{j=0}^{\infty}\phi^{j}W_{-j}$$

Consequently, the dispersion is

$$\mathrm{disp}(\underline{m}'\underline{X}_n - X_{n+k}) = 1 + |\phi+\theta|^\alpha \frac{1-|\phi|^{\alpha(k-1)}}{1-|\phi|^\alpha} + \sum_{j=1}^{n} |c_j + \theta c_{j-1}|^\alpha$$

$$+ \frac{|\phi+\theta|^\alpha}{1-|\phi|^\alpha} |c_n|^\alpha \tag{3.4}$$

where $c_j = \sum_{i=0}^{j} m_i \phi^{j-i}$, $j \geq 0$ (and $m_j = c_j - \phi c_{j-1}, j \geq 1$). It suffices, then, to minimize

$$h(\underline{c}) = \sum_{j=1}^{n} |c_j + \theta c_{j-1}|^\alpha + \frac{|\phi+\theta|^\alpha}{1-|\phi|^\alpha} |c_n|^\alpha \tag{3.5}$$

and this will be done recursively, minimizing first with respect to c_n, then c_{n-1} and so on.

Assume first that $|\phi+\theta|^\alpha \leq 1-|\phi|^\alpha$. By Lemma 3.1, for fixed $c_{n-1}, \ldots, c_1$, $h(\underline{c})$ is minimized by choosing $c_n = -\theta c_{n-1}$. Under this condition (3.5) becomes

$$\min_{c_n} h(\underline{c}) = \sum_{j=1}^{n-1} |c_j + \theta c_{j-1}|^\alpha + |\theta|^\alpha \frac{|\phi+\theta|^\alpha}{1-|\phi|^\alpha} |c_{n-1}|^\alpha .$$

Since $|\theta| < 1$ (and hence $|\theta|^\alpha |\phi+\theta|^\alpha < 1-|\phi|^\alpha$), then $h(\underline{c})$ is minimized further by choosing $c_{n-1} = -\theta c_{n-2}$, again using Lemma 3.1. The resulting value for $h(\underline{c})$ will have a similar form so that continuing recursively, we can choose $c_j = -\theta c_{j-1}$, $1 \leq j \leq n$. Since $c_0 = m_0 = -\phi^{k-1}$, then $c_j = -(-\theta)^j \phi^{k-1}$ and the minimizing vector $\underline{m}$ is $\underline{a}$, as given in (3.2). The minimum value of $\mathrm{disp}(X_{n+k} - \underline{a}'X_n)$ is

$$1+|\phi+\theta|^{\alpha}\ \frac{1-|\phi|^{\alpha(\ -1)}}{1-|\phi|^{\alpha}} + \min h(\underline{c})$$

$$= 1+|\phi+\theta|^{\alpha}\ \frac{1-|\phi|^{\alpha(k-1)}}{1-|\phi|^{\alpha}} + |\theta|^{n\alpha}|\phi|^{\alpha(k-1)}\ \frac{|\phi+\theta|^{\alpha}}{1-|\theta|^{\alpha}}\ .$$

The second case is where $|\phi+\theta|^{\alpha} > 1-|\phi|^{\alpha}$. The argument is the same, except that first we choose $c_n=0$, according to Lemma 3.1, to minimize (3.5). In this case,

$$\min_{c_n} h(\underline{c}) = \sum_{j=1}^{n-1} |c_j+\theta c_{j-1}|^{\alpha} + |\theta|^{\alpha}\ |c_{n-1}|^{\alpha}. \tag{3.6}$$

Since $|\theta|<1$, then (3.5) is further minimized by setting $c_j = -\theta c_{j-1}$, $1\le j\le n-1$, as done previously. Again using $c_0 = -\phi^{k-1}$ and $m_j = c_j-\phi c_{j-1}$, we have $\underline{a}$, as in (3.2), is the minimizing vector $\underline{m}$. The minimum error dispersion is

$$1+|\phi+\theta|^{\alpha}\ \frac{1-|\phi|^{\alpha(k-1)}}{1-|\phi|^{\alpha}} + \min h(\underline{c})$$

$$= 1+|\phi+\theta|^{\alpha}\ \frac{1-|\phi|^{\alpha(k-1)}}{1-|\phi|^{\alpha}} + |\theta|^{n\alpha}|\phi|^{\alpha(k-1)}.$$

Finally, the choice of $\underline{a}$ is unique, according to the lemma, except when $|\theta+\phi|^{\alpha} = 1-|\phi|^{\alpha}$. In this case, the final coefficient a_n may be chosen in either of the two ways given in (3.2). #

Remarks

1. Except possibly for the final term, this predictor is the same as the "truncated" version of X_{n+k}. That is, if we write $X_{n+k} = W_{n+k} +\psi_1 X_{n+k-1}+\psi_2 X_{n+k-2}+\ldots$, for some sequence $\{\psi_j\}$, then the truncated

predictor presumes all unobserved values are zero, $X^*_{n+k} = \psi_k X_n + \ldots + \psi_{k+n-1} X_1$.

2. The special case $\theta=0$ is the AR(1) process already treated in Section 2. We find as before that for $n \geq 1$, $k \geq 1$, the minimum dispersion predictor of X_{n+k} is $\hat{X}_{n+k} = \phi^k X_n$ and the error dispersion is

$$\frac{1-|\phi|^{\alpha k}}{1-|\phi|^{\alpha}}.$$

A more interesting special case is the process MA(1) obtained when $\phi=0$. For this case, $\hat{X}_{n+1} = -\sum_{j=0}^{n} (-\theta)^j X_{n+1-j}$, $\hat{X}_{n+k} = 0$ for $k \geq 2$. The error dispersion for $\hat{X}_{n+1}$ is $1+|\theta|^{(n+1)\alpha}$ and for $\hat{X}_{n+k} (k \geq 2)$, it is $1+|\theta|^{\alpha}$.

3. Although the prediction is not necessarily unique, it can be defined in such a way as to correspond to a linear mapping $Y \to \hat{Y}$ or span $\{X_1, X_2, \ldots\}$. To see this, we need only to observe that for each $j \geq 1$

$$X_j = W_j + (\phi+\theta) \sum_{i=1}^{j-1} \phi^{i-1} W_{j-i} + \phi^{j-1}(\phi+\theta) W_0^*$$

where $W_0^* = \sum_{i=0}^{\infty} \phi^i W_{-i}$, and apply Theorem 3.5 from Part I. In particular, this allows us to write

$$\begin{aligned} \hat{X}_{n+k} &= \hat{W}_{n+k} + \theta \hat{W}_{n+k-1} + \phi \hat{X}_{n+k-1} \\ &= \phi X_{n+k-1} \\ &= \phi^{k-1} \hat{X}_{n+1} \end{aligned}$$

and this agrees with Theorem 3.2.

4. We can also obtain predictors recursively using $\hat{X}_{j(k)} = P_k X_j$, the predictor for X_j, based on $X_k, \ldots, X_i$. The simple formula is

$$X_{n+1(n)} = \phi X_n + \theta(\hat{X}_{n(n-1)} - X_n),$$

with

$$\hat{X}_{2(1)} = \begin{cases} (\phi+\theta)X_1 & \text{if } |\phi+\theta|^\alpha \le 1-|\phi|^\alpha \\ \phi X_1 & \text{otherwise} \end{cases}$$

As we shall see, the linearity property and the recursion formula will extend at least partially to the ARMA(1,1) model with $\alpha>1$. They do not extend, however, to more general ARMA models.

<u>Theorem 3.3</u> For the ARMA(1,1) process (3.1) with $\alpha>1$, there is a unique minimum dispersion linear predictor $\hat{X}_{n+k} = \underline{a}'\underline{X}_n$ for X_{n+k}. The vector $\underline{a}$ is given by

$$a_j = \phi^{k-1}(-\theta)^{j-1}\,\frac{(\phi+\theta)(1-\eta+\xi)-\xi\eta^{n-j}(\eta\phi+\theta)}{1-\eta+\xi(1-\eta^n)}, \quad 1 \le j \le n \tag{3.7}$$

where $\eta = |\theta|^{\alpha/\alpha-1}$ and $\xi = \left(\frac{|\phi+\theta|^\alpha}{1-|\phi|^\alpha}\right)^{1/\alpha-1}$. The minimum error dispersion is

$$\text{disp}(X_{n+k}-\underline{a}'\underline{X}_n) = 1+\xi^{\alpha-1}(1-|\phi|^{\alpha(k-1)}) + \left(\frac{\xi\eta^n(1-\eta)}{1-\eta+\xi(1-\eta^n)}\right)^{\alpha-1}.$$

<u>Proof</u>: As in the proof of Theorem 3.2, we minimize

$$h(c) = \sum_{j=1}^{n} |c_j + \theta c_{j-1}|^\alpha + \xi^{\alpha-1}|c_n|^\alpha, \tag{3.8}$$

first with respect to c_n, then c_{n-1} and so on, subject to the condition $c_0 = -\phi^{k-1}$.

Using Lemma (3.1) (now with $\alpha>1$),

$$c_n = -\theta c_{n-1}(1+\xi)^{-1}$$

with corresponding value for (3.8)

$$\min_{c_n} h(\underline{c}) = \sum_{j=1}^{n-1} |c_j+\theta c_{j-1}|^\alpha + \left(\frac{\eta\xi}{1+\xi}\right)^{\alpha-1} |c_{n-1}|^\alpha .$$

This is further minimized when

$$c_{n-1} = -\theta c_{n-2}\left(1+\frac{\eta\xi}{1+\xi}\right)^{-1}$$

and then

$$\min_{c_n,c_{n-1}} h(\underline{c}) = \sum_{j=1}^{n-2} |c_j+\theta c_{j-1}|^\alpha + \left(\frac{\eta^2\xi}{1+\xi+\eta\xi}\right)^{\alpha-1} |c_{n-2}|^\alpha .$$

Continuing the stepwise minimization we find that

$$c_j = -\theta c_{j-1} \frac{1-\eta+\xi(1-\eta^{n-j})}{1-\eta+\xi(1-\eta^{n-j+1})}$$

Since $c_0 = -\phi^{k-1}$, we deduce that

$$c_j = -\phi^{k-1}(-\theta)^j \frac{1-\eta+\xi(1-\eta^{n-j})}{1-\eta+\xi(1-\eta^{n})} . \tag{3.9}$$

From this, and the relations $a_j = c_j - \phi c_{j-1}$, we get $\underline{a}$ satisfying (3.7) as the unique solution to minimizing $\text{disp}(X_{n+k} - \underline{a}'\underline{X}_n)$. The minimum error dispersion, from the expressions (3.4), (3.8) and (3.9), is

$$\operatorname{disp}(X_{n+k}-\underline{a}'\underline{X}_n) = 1+\xi^{\alpha-1}(1-|\phi|^{\alpha(k-1)}) + \min h(\underline{c})$$

$$= 1+\xi^{\alpha-1}(1-|\phi|^{\alpha(k-1)}) + \left(\frac{\xi\eta^n(1-\eta)}{1-\eta+\xi(1-\eta^n)}\right)^{\alpha-1} . \qquad \#$$

Remarks

1. For the special case of an AR(1) process, $\theta=\eta=0$, $a_1=\phi^k$ and $a_j=0$, as found earlier in Corollary 2.2. The minimum error dispersion is $\frac{1-|\phi|^{\alpha k}}{1-|\phi|^{\alpha}}$.

2. For the MA(1) case, $\phi=0$ and $\eta=\xi=|\theta|^{\alpha/\alpha-1}$. To predict X_{n+1}, $a_j = -(-\theta)^j \frac{1-\eta^{n+1-j}}{1-\eta^{n+1}}$, $1\leq j\leq n$, and the minimum error dispersion is $1 + \left(\frac{\eta^{n+1}(1-\eta)}{1-\eta^n}\right)^{\alpha-1}$. For $k\geq 2$, $\hat{X}_{n+k}=0$ and the error dispersion is $1+|\theta|^{\alpha}$.

3. Here again we get a partial linearity property for the operator, namely that for $Y = \ell_1 X_{n+1}+\ldots+\ell_k X_{n+k}$,

$$\hat{Y} = \ell_1\hat{X}_{n+1}+\ldots+\ell_k\hat{X}_{n+k} = (\ell_1+\phi\ell_2+\ldots+\phi^{k-1}\ell_k)\hat{X}_{n+1}.$$

This can be determined by minimizing $h(\underline{c})$ in (3.8), now subject to $C_0 = -(\ell_1+\phi\ell_2+\ldots+\phi^{k-1}\ell_k)$.

4. As in the $\alpha\leq 1$ case, the minimum dispersion predictor can be obtained recursively. Let $\hat{X}_{j(k)} = P_k X_j$. Then by a straightforward

calculation, $\hat{X}_{n+1(n)} = \phi X_n + \theta \frac{\nu_{n-1}}{\nu_n}(\hat{X}_{n(n-1)} - X_n)$, where $\nu_n = 1+\eta+\xi(1-\eta^n)$, $\hat{X}_{1(0)} = 0$.

5. Minimizing (3.8) with $\alpha=2$ gives the least squares predictor for X_{n+k}, which is $\tilde{X}_{n+k} = b'\underline{X}_n$ where

$$b_j = \phi^{k-1}(-\theta)^{j-1}(\phi+\theta)\left(\frac{1-\theta\rho\theta^{2(n-j)}}{1-\rho^2\theta^{2n}}\right), \quad \rho = \frac{\phi+\theta}{1+\phi\theta},$$

and the error dispersion of this predictor, for any α, is

$$1 + |\phi+\theta|^\alpha \frac{1-|\phi|^{\alpha(k-1)}}{1-|\phi|^\alpha} + \left|\frac{\theta^n}{1-\rho^2\theta^{2n}}\right|^\alpha \left(|\rho^2(1-\theta^2)|^\alpha \frac{1-|\theta|^{n\alpha}}{1-|\theta|^\alpha} + \frac{|(\phi+\theta)(1-\rho^2)|^\alpha}{1-|\phi|^\alpha}\right).$$

The least squares predictor $\tilde{X}_{n+1(n)}$ is recursively calculated by $\tilde{X}_{n+1(n)} = \phi X_n + \theta \frac{1-\rho^2\theta^{2(n-1)}}{1-\rho^2\theta^{2n}}(\tilde{X}_{n(n-1)} - X_n)$, $\tilde{X}_{1(0)} = 0$. (See also Brockwell and Davis (1983) for a general discussion of least squares prediction).

4. Prediction for the MA(q) and ARMA(p,q) Models

Assume the process $\{X_n\}$ satisfies $X_n = W_n+\theta_1 W_{n-1}+\ldots+\theta_q W_{n-q}$ where $(1+\theta_1 z+\ldots+\theta_q z^q) \neq 0$ for complex $|z|\leq 1$. In order to predict X_{n+1}, we need to minimize

$$\operatorname{disp}(X_{n+1}-\underline{a}'\underline{X}_n) = 1 + \sum_{j=1}^{n+q} |a_j+a_{j-1}\theta_1+\ldots+a_{j-q}\theta_q|^\alpha \tag{4.1}$$

where $a_0=-1$ and $a_j=0$ for $j<0$ or $j>n$. According to Part I, Theorem 3.4, when $\alpha\leq 1$ it suffices to consider only $\underline{a}\varepsilon R^n$ which satisfy

$$a_j+a_{j-1}\theta_1+\ldots+a_{j-q}\theta_q = 0 \tag{4.2}$$

for at least n of the n+q equations, $1\leq j\leq n+q$. The set of choices is thus limited to $\binom{n+q}{q}$ possibilities. In Theorem 3.2 we have already established which choice is optimal for the MA(1) model. Exactly one choice was the best for all values of θ_1 in the parameter space, $|\theta_1|<1$. If $q>1$, however, the optimal formula depends on the particular region of the parameter space. We look specifically at the MA(2) model.

Lemma 4.1 Suppose $\{X_n\}$ is an MA(2) process with $\alpha\leq 1$. Define z_1,z_2 to be the solutions to $(z^2+\theta_1 z+\theta_2) = 0$, and

$$S_j = \begin{cases} \dfrac{z_1^j-z_2^j}{z_1-z_2} & \text{if } z_1\neq z_2 \\[2ex] j z_1^{j-1} & \text{if } z_1=z_2 . \end{cases}$$

Then the minimum dispersion predictor for X_{n+1} lies in the set of the $\binom{n+2}{2}$ choices for $\underline{a}'\underline{X}_n$ where $1\leq j_1<j_2\leq n+2$ and

$$a_j = \begin{cases} -S_{j+1} & \text{if } 1\leq j<j_1 \\ -\theta_2^{\,j-j_1+1}\,\dfrac{S_{j_2-j-1}S_{j_1}}{S_{j_2-j_1}} & \text{if } j_1\leq j<j_2 \\ 0 & \text{if } j_2\leq j\leq n \end{cases}$$

The dispersion of the prediction error is

$$\text{disp}(X_{n+1}-\underline{a}'\underline{X}_n) = 1 + \left|\frac{S_{j_2}}{S_{j_2-j_1}}\right|^{\alpha} + \left|\theta_2^{\,j_2-j_1}\frac{S_{j_1}}{S_{j_2-j_1}}\right| . \qquad (4.3)$$

Proof: We recognize that the S_j's can be determined recursively by $S_{j+2}+\theta_1S_{j+1}+\theta_2S_j = 0$ and that in fact $S_{j+1}S_{k+1}-\theta_2S_jS_k = S_{j+k+1}$. From these we can easily verify that (using $a_0=-1$, $a_{-1}=a_{n+1}=a_{n+2} = 0$ and fixing j_1,j_2)

$$a_j+\theta_1a_{j-1}+\theta_2a_{j-2} = \begin{cases} 0 & \text{for } j\neq j_1,\ j\neq j_2 \\ -\dfrac{S_{j_2}}{S_{j_2-j_1}} & \text{for } j=j_1 \\ -\theta_2^{\,j_2-j_1}\dfrac{S_{j_1}}{S_{j_2-j_1}} & \text{for } j=j_2 . \end{cases}$$

Thus for each of the $\binom{n+2}{n}$ choices of j_1 and j_2, $\underline{a}$ satisfies n of the n+2 equations in (4.2). It follows that a minimum predictor is of this form. The error dispersion is

$$\mathrm{disp}(X_{n+1}-\underline{a}'\underline{X}_n) = 1 + \sum_{j=1}^{n+2} |a_j+\theta_1 a_{j-1}+\theta_2 a_{j-2}|^\alpha$$

$$= 1 + \left|\frac{S_{j_2}}{S_{j_2-j_1}}\right|^\alpha + \left|\theta_2^{\,j_2-j_1}\frac{S_{j_1}}{S_{j_2-j_1}}\right|^\alpha .$$

To actually determine the predictor, we need to minimize (4.3) over the $\binom{n+2}{n}$ possible choices of j_1,j_2. #

For a more general MA(q) process, with $\alpha>1$, the minimum dispersion predictor is obtained by solving for $\underline{a}\varepsilon R^n$ to satisfy (using $[x]^{\alpha-1} = \mathrm{sgn}(x)|x|^{\alpha-1}$)

$$[a_j+\theta_1 a_{j-1}+\ldots+\theta_q a_{j-q}]^{\alpha-1}+\theta_1[a_{j+1}+\theta_1 a_j+\ldots+\theta_q a_{j-q+1}]^{\alpha-1}+$$

$$\ldots+\theta_q[a_{j+q}+\ldots+\theta_q a_j]^{\alpha-1} = 0 \qquad 1\le j\le n .$$

This can be accomplished recursively in the following manner: Let

$$\Pi = \begin{bmatrix} 1 & \theta_1 & \theta_2 & \cdots & \theta_q & 0 & \cdots & 0 \\ 0 & 1 & \theta_1 & \cdots & & \theta_q & \cdots & 0 \\ \vdots & & \ddots & & & & & \\ & & & 1 & \theta_1 & & \cdots & \theta_q \end{bmatrix} \quad (n\times n+q) \text{ matrix}$$

and

$$\underline{\rho} = [\theta_1\ \theta_2\ \ldots\ \theta_q\quad 0\ \ldots\ 0] \qquad (n+q) \text{ vector.}$$

Set $\underline{a}_0 = (\Pi\,\Pi')^{-1}\Pi\underline{\rho}$. ($\underline{a}_0'\underline{X}_n$ is the least squares predictor.) Next define $\ell_j(\underline{a}) = [\underline{a}'\underline{\Pi}_j-\rho_j]^{\alpha-1} = [-\theta_j+a_1\theta_{j-1}+\ldots+a_q\theta_{j-q}]^{\alpha-1}$ $1\le j\le n+2$.

The recursion is then given by

$$\underline{a}_{k+1} = \underline{a}_k - (\Pi\ \Pi')^{-1}\Pi\underline{\ell}(\underline{a}_k).$$

The ARMA (1,q) process can be handled similarly to the MA(q) process. Let $\pi_j = \phi^j+\phi^{j-1}\theta_1+\ldots+\phi\theta_{j-1}+\theta_j$. The process can be expressed by

$$X_n = W_n+\pi_1 W_{n-1}+\ldots+\pi_{n+q-2}W_{2-q}+\frac{\pi_{n+q-1}}{(1-|\phi|^\alpha)^{1/\alpha}}W^*,\ n\geq 1$$

where $W^* = (1-|\phi|^\alpha)^{1/\alpha}\sum_{j=0}^{\infty}\phi^j W_{1-q-j} \overset{d}{=} W_{1-q}$ and is independent of W_{2-q}, W_{3-q}, ... To predict X_{n+1} we need to minimize

$$\begin{aligned}\operatorname{disp}(X_{n+1}-\underline{a}'X_n) &= 1+\sum_{j=1}^{n+q-1}|a_j+\pi_1 a_{j-1}+\ldots+\pi_j a_0|^\alpha \\ &\quad+\frac{|a_{n+q}+\ldots+\pi_{n+q}a_0|^\alpha}{1-|\phi|^\alpha} \\ &= 1+\sum_{j=1}^{n+q-1}|c_j+\theta_1 c_{j-1}+\ldots+\theta_q c_{j-q}|^\alpha+\frac{|\pi_q c_n|^\alpha}{1-|\phi|^\alpha},\end{aligned}$$

where $c_j = \sum_{k=0}^{j}\phi^k z_{j-k}, a_j = c_j-\phi c_{j-1}$, $j\geq 0$ and $c_j=0$, $j<0$. Except for the last term this (as a function of $\underline{c}$) is similar to (4.1). The minimization is thus done with respect to $\underline{c}$ and then $\underline{a}$ is obtained from $\underline{c}$.

The ARMA(1,q) minimization involves a finite sum. This is not true for the more general ARMA(p,q) process. By defining the c_j's to satisfy $a_j = c_j-\phi_1 c_{j-1}-\ldots-\phi_p c_{j-p}$, $j\geq 0$, $c_j=0$ for $j<0$, then to predict X_{n+1} in an ARMA(p,q) process, we minimize

$$\text{disp}(X_{n+1}-\underline{a}'\underline{X}_n) = 1 + \sum_{j=1}^{\infty} |c_j+\theta_1 c_{j-1}+\ldots+\theta_q c_{j-q}|^{\alpha}$$

$$= 1 + \sum_{j=1}^{n} |c_j+\theta_1 c_{j-1}+\ldots+\theta_q c_{j-q}|^{\alpha}$$

$$+ \sum_{j=1}^{\infty} |\sigma_{j1}c_n+\ldots+\sigma_{jn}c_1|^{\alpha}$$

where σ_{ji} are chosen so that $c_{n+j}+\theta_1 c_{n+j-1}+\ldots+\theta_q c_{n+j-q} = \sigma_{j1}c_n+ \ldots+\sigma_{jn}c_1$ $(n>\max(p,q+1))$. The sum can be truncated after an appropriate number of terms to facilitate the minimization. Alternatively, we write $X_n = W_n + \sum_{j=1}^{\infty} \psi_j X_{n-j}$ for some sequence $\{\psi_j\}$, and predict X_{n+1} with $X^*_{n+1} = \sum_{j=1}^{n} \psi_j X_{n+1-j}$. This is the so-called truncation predictor which is often used in place of least squares when the model is of large order.

5. Numerical Comparison of Minimum Dispersion and Least Squares Predictors

In this section we compare the minimum dispersion predictor (MDP) $\hat{X}_{n+1}$ with the least squares predictor (LSP) $\tilde{X}_{n+1}$, in the special case of the ARMA(1,1) process. We recall here the error dispersions of these predictors.

For MDP,

$$\gamma_{MD} = \operatorname{disp}(\hat{X}_{n+1}-X_{n+1}) = \begin{cases} 1 + |\theta|^{n\alpha}\min(1,\tau) & \text{if } \alpha\le 1 \\ 1 + |\theta|^{n\alpha}\tau\left(\dfrac{1-\eta}{1-\eta+\tau^{1/\alpha-1}(1-\eta^n)}\right) & \text{if } \alpha>1 \end{cases}$$

where

$$\tau = \frac{|\phi+\theta|^\alpha}{1-|\phi|^\alpha}, \quad \eta = |\theta|^{\alpha/\alpha-1}.$$

For LSP, $\gamma_{LS} = \operatorname{disp}(\tilde{X}_{n+1}-X_{n+1}) = 1 + |\theta|^{n\alpha}\left[\tau\left(\dfrac{1-\rho^2}{1-\rho^2\theta^{2n}}\right)^\alpha\right.$

$\left.+\sigma\left(\dfrac{\rho^2(1-\theta^2)}{1-\rho^2\theta^{2n}}\right)^\alpha\right]$ where $\sigma = \dfrac{1-|\theta|^{n\alpha}}{1-|\theta|^\alpha}$ and $\rho = \dfrac{\phi+\theta}{1+\phi\theta}$ and τ is as above.

From these expressions it is quite clear that whenever $|\theta|^{n\alpha}\tau$ is small, then $\gamma_{LS}-\gamma_{MD}$ will be also. Thus LSP, which has the larger dispersion, will be nearly equivalent to MDP when either

i) $|\theta|$ is small

ii) $|\phi+\theta|$ is small

iii) n is large

iv) α is large.

Table 5.1 gives a comparison of dispersions for several models and we see that for $\alpha=1.5$ there is little difference in the two dispersions for these models. With $\alpha=0.5$, however, the difference $\gamma_{LS}-\gamma_{MD}$ is more noticeable, especially for the larger value of θ. For $\alpha=1.0$, the difference is substantial only for the higher value of θ and the smaller sample size.

The general conclusion we can make then, is that for ARMA(1,1) minimum dispersion is preferable to least squares when $\alpha\leq 1$ and especially when there is a large moving average parameter and the number of observations is relatively small. The fact that for $\alpha\leq 1$, the minimum dispersion predictor is easy to calculate makes it more appealing.

In order to make a more careful comparison, we provide the results of a simulation for three of the models in which γ_{LS} was somewhat larger than γ_{MD}. Tables 5.2, 5.3 and 5.4 provide the coefficients for the predictors, $\hat{X}_{n+1} = a_1X_n+\ldots+a_nX_1$ and $\tilde{X}_{n+1} = b_1X_n+\ldots+b_nX_1$, for each of these models. In each simulation, 2500 or 1000 independent series were generated. The noise terms were simulated by $W_j = [\tan(\pi[U_j]^{1/\alpha}/2)]^{\alpha}$ where $\{U_j\}$ were independent psuedo-uniform random variables.

Let $\hat{F}_{MD}$ and $\hat{F}_{LS}$ be the empirical distributions for the absolute MDP errors (MDE) and the absolute LSP errors (LSE). Each table gives the value $\ln x$ vs. $-\ln(1-\hat{F}_{MD}(x))$ and vs. $-\ln(1-\hat{F}_{LS}(x))$. If graphed, these would be approximately straight lines, because $1-F_{MD}(x) \sim \gamma_{MD}x^{-\alpha}L(x)$ and $1-F_{LS}(x) \sim \gamma_{LS}x^{-\alpha}L(x)$ where F_{MD} and F_{LS} are the true absolute error distributions and $L(x)$ is a slowly varying function (in this example $L(x)\to$ constant). The difference between the two lines estimates $\ln(\gamma_{LS}) - \ln(\gamma_{MD})$. We remark that in the model with $\alpha=1.0$, the data are

Cauchy. For Cauchy data (and whenever the data are symmetric stable), $1-F_{MS}(x) < 1-F_{LS}(x)$ for all $x>0$.

Since both $\hat{X}_{n+1}$ and $\tilde{X}_{n+1}$ were obtained for each series, we can make a pairwise comparison. Each of the three tables includes a frequency table for $\ln\left(\frac{|\tilde{X}_{n+1}-X_{n-1}|}{|\hat{X}_{n+1}-X_{n+1}|}\right)$. Also included is the percentage of series for which the least squares error (LSE) exceeded the minimum dispersion error (MSE) in absolute value. For these models, at least, LSE exceeded MSE more often than not.

Table 5.1 A Comparison of Error Dispersion in

Predicting X_{n+1} from an ARMA(1,1) Process

Size	α	ϕ	θ	MDP DISP	LSP DISP
10	.50	.7000	.5000	1.03125	1.17558
10	.50	.3000	.5000	1.03125	1.10667
10	.50	0.0000	.5000	1.02210	1.06389
10	.50	-.3000	.5000	1.03090	1.05109
10	.50	-.7000	.5000	1.03125	1.10895
10	.50	.7000	.8000	1.32768	2.88886
10	.50	.3000	.8000	1.32768	2.46812
10	.50	0.0000	.8000	1.29309	2.18187
10	.50	-.3000	.8000	1.32768	2.21258
10	.50	-.7000	.8000	1.32768	1.90262
10	1.00	.7000	.5000	1.00098	1.00198
10	1.00	.3000	.5000	1.00098	1.00128
10	1.00	0.0000	.5000	1.00049	1.00073
10	1.00	-.3000	.5000	1.00028	1.00034
10	1.00	-.7000	.5000	1.00065	1.00073
10	1.00	.7000	.8000	1.10737	1.20216
10	1.00	.3000	.8000	1.10737	1.17329
10	1.00	0.0000	.8000	1.08590	1.14239
10	1.00	-.3000	.8000	1.07670	1.11876
10	1.00	-.7000	.8000	1.03579	1.04288
10	1.50	.7000	.5000	1.00000	1.00000
10	1.50	.3000	.5000	1.00001	1.00001
10	1.50	0.0000	.5000	1.00001	1.00001
10	1.50	-.3000	.5000	1.00000	1.00000
10	1.50	-.7000	.5000	1.00001	1.00001
10	1.50	.7000	.8000	1.00009	1.00010
10	1.50	.3000	.8000	1.00202	1.00218
10	1.50	0.0000	.8000	1.00600	1.00653
10	1.50	-.3000	.8000	1.00797	1.00870
10	1.50	-.7000	.8000	1.00262	1.00274
25	.50	.7000	.5000	1.00017	1.00098
25	.50	.3000	.5000	1.00017	1.00060
25	.50	0.0000	.5000	1.00012	1.00036
25	.50	-.3000	.5000	1.00017	1.00029
25	.50	-.7000	.5000	1.00017	1.00061
25	.50	.7000	.8000	1.06146	1.44183
25	.50	.3000	.8000	1.06146	1.35663
25	.50	0.0000	.8000	1.05498	1.29527
25	.50	-.3000	.8000	1.06146	1.28806
25	.50	-.7000	.8000	1.06146	1.19039
25	1.00	.7000	.5000	1.00000	1.00000
25	1.00	.3000	.5000	1.00000	1.00000
25	1.00	0.0000	.5000	1.00000	1.00000
25	1.00	-.3000	.5000	1.00000	1.00000
25	1.00	-.7000	.5000	1.00000	1.00000

Table 5.1 (Continued)

Size	α	ϕ	θ	MDP DISP	LSP DISP
25	1.00	.7000	.8000	1.00378	1.00769
25	1.00	.3000	.8000	1.00378	1.00660
25	1.00	0.0000	.8000	1.00302	1.00542
25	1.00	-.3000	.8000	1.00270	1.00446
25	1.00	-.7000	.8000	1.00126	1.00154
25	1.50	.7000	.5000	1.00000	1.00000
25	1.50	.3000	.5000	1.00000	1.00000
25	1.50	0.0000	.5000	1.00000	1.00000
25	1.50	-.3000	.5000	1.00000	1.00000
25	1.50	-.7000	.5000	1.00000	1.00000
25	1.50	.7000	.8000	1.00000	1.00000
25	1.50	.3000	.8000	1.00001	1.00001
25	1.50	0.0000	.8000	1.00004	1.00004
25	1.50	-.3000	.8000	1.00005	1.00006
25	1.50	-.7000	.8000	1.00002	1.00002

Table 5.2

2500 Series Of Size 10, Predicting Observation 11
α = 0.50 ϕ = .000 θ = .800
MDP Error Dispersion Is 1.29309
LSP Error Dispersion Is 2.18187

Coefficients	MDP	LSP
j	a_j	b_j
1	.80000	.79665
2	-.64000	-.63314
3	.51200	.50129
4	-.40960	-.39450
5	.32768	.30743
6	-.26214	-.23574
7	.20972	.17583
8	-.16777	-.12471
9	.13422	.07983
10	-.10737	-.03894

Log Cumulative Tail Distribution For Errors

lnx	$-\ln(1-\hat{F}_{MD}(x))$	$-\ln(1-\hat{F}_{LS}(x))$
0.00	.1999	.0934
.20	.2337	.1112
.40	.2687	.1319
.60	.3153	.1550
.80	.3613	.1873
1.00	.4210	.2226
1.20	.4742	.2593
1.40	.5447	.3054
1.60	.6221	.3510
1.80	.6988	.4017
2.00	.7696	.4532
2.20	.8468	.5189
2.40	.9314	.5770
2.60	1.0217	.6394
2.80	1.1062	.7060
3.00	1.2066	.7739
3.20	1.2845	.8496
3.40	1.3799	.9153
3.60	1.4908	.9921
3.80	1.5896	1.0871
4.00	1.6671	1.1533
4.20	1.7373	1.2174
4.40	1.8351	1.3034
4.60	1.9296	1.3847
4.80	2.0129	1.4490
5.00	2.1103	1.5437

Table 5.2 (Continued)

Frequency Table for log(LSE/MDE)

Interval	Frequency
$-\infty$, -1.00	361
-1.00 , - .60	143
- .60 , - .12	235
- .12 , .00	197
.00 , .12	308
.12 , .60	274
.60 , 1.00	157
1.00 , ∞	825

LSE Exceeded MDE 62.56% of the Time

Table 5.3

2500 Series Of Size 10, Predicting Observation 11
α = 1.00 ϕ = .700 θ = .800
MDP Error Dispersion Is 1.10737
LSP Error Dispersion Is 1.20216

Coefficients	MDP	LSP
j	a_j	b_j
1	1.50000	1.49515
2	-1.20000	-1.18667
3	.96000	.93752
4	- .76800	- .73524
5	.61440	.56973
6	- .49152	- .43270
7	.39322	.31731
8	- .31457	- .21778
9	.25166	.12914
10	- .09395	- .04696

Log Cumulative Tail Distribution For Errors

lnx	$-\ln(1-\hat{F}_{MD}(x))$	$-\ln(1-\hat{F}_{LS}(x))$
0.00	.6440	.5855
.20	.7722	.7248
.40	.9243	.8685
.60	1.0883	1.0106
.80	1.2888	1.1738
1.00	1.5105	1.3720
1.20	1.7016	1.5799
1.40	1.8708	1.7510
1.60	2.0875	1.9296
1.80	2.2595	2.1136
2.00	2.4865	2.3351
2.20	2.7001	2.5158
2.40	2.8268	2.6536
2.60	3.0619	2.8682
2.80	3.2289	3.0791
3.00	3.4420	3.2702
3.20	3.6194	3.5066
3.40	3.8167	3.6497
3.60	3.9528	3.7297
3.80	4.0628	3.9120
4.00	4.3901	4.2687
4.20	4.7330	4.5282
4.40	4.9908	4.7795
4.60	5.2591	5.0515
4.80	5.2591	5.0515
5.00	5.2591	5.0515

Table 5.3 (Continued)

Frequency Table for log(LSE/MDE)

Interval	Frequency
$-\infty$, -1.00	195
-1.00 , - .60	130
.60 , - .12	421
.12 , .00	441
.00 , .12	449
.12 , .60	456
.60 , 1.00	121
1.00 , ∞	287

LSE Exceeded MDE 52.52% of the Time

Table 5.4

1000 Series Of Size 25, Predicting Observation 26
α = 0.50 ϕ = .700 θ = .800
MDP Error Dispersion Is 1.06146
LSP Error Dispersion Is 1.44183

Coefficients	MDP	LSP
j	a_j	b_j
1	1.50000	1.49999
2	-1.20000	-1.19998
3	.96000	.95997
4	-.76800	-.76796
5	.61440	.61435
6	-.49152	-.49145
7	.39322	.39312
8	-.31457	-.31445
9	.25166	.25151
10	-.20133	-.20114
11	.16106	.16082
12	-.12885	-.12855
13	.10308	.10271
14	-.08246	-.08200
15	.06597	.06539
16	-.05278	-.05205
17	.04222	.04131
18	-.03378	-.03263
19	.02702	.02559
20	-.02162	-.01983
21	.01729	.01506
22	-.01384	-.01105
23	.01107	.00758
24	-.00885	-.00450
25	.00331	.00163

Log Cumulative Tail Distribution For Errors

$\ln x$	$-\ln(1-\hat{F}_{MD}(x))$	$-\ln(1-\hat{F}_{LS}(x))$
0.00	.2485	.1744
.20	.3133	.2046
.40	.3725	.2370
.60	.4170	.2758
.80	.4878	.3313
1.00	.5516	.3667
1.20	.6070	.4277
1.40	.6911	.4829
1.60	.7508	.5534

Table 5.4 (Continued)

Log Cumulative Tail Distribution For Errors (Continued)

1.80	.8142	.6015
2.00	.8892	.6655
2.20	.9597	.7340
2.40	1.0441	.8074
2.60	1.1363	.9039
2.80	1.2483	1.0189
3.00	1.3626	1.1270
3.20	1.4567	1.2174
3.40	1.5233	1.2874
3.60	1.5896	1.3704
3.80	1.6766	1.4355
4.00	1.7487	1.4917
4.20	1.8452	1.5654
4.40	1.9310	1.6766
4.60	2.0174	1.7487
4.80	2.0956	1.8202
5.00	2.1982	1.9173

Frequency Table for log(LSE/MDE)

Interval	Frequency
$-\infty$, -1.00	51
-1.00 , - .60	33
- .60 , - .12	100
- .12 , .00	205
.00 , .12	243
.12 , .60	108
.60 , 1.00	56
1.00 , ∞	148

LSE Exceeded MDE 61.10% of the Time

PART III: PRODUCTS OF INDEPENDENT RANDOM VARIABLES AND DOMAINS OF ATTRACTION

Summary. We consider sufficient conditions for the distribution of the product of two independent random variables to be in the domain of attraction of a stable law. We also consider conditions for the component wise product of two independent pairs of random variables to be in a bivariate domain of attraction. Included are two results from the literature concerning regular variation of the tail distribution of such products.

1. Introduction

In this paper we shall have two settings. For the first, $X \sim F$ and $Y \sim G$ will be independent random variables. Letting H be the distribution of XY, we ask two questions: under what conditions on F and G are the tails of H regularly varying and under what conditions is H in the domain of attraction of a stable law? In the second setting, $(X_1,X_2) \sim F$ and $(Y_1,Y_2) \sim G$ are independent pairs of random variables. Let H be the distribution of (X_1Y_1,X_2Y_2) and we ask, when will H be in the domain of attraction of a bivariate stable law?

Before we proceed, we recall some definitions. A positive function U on $(0,\infty)$ is regularly varying with exponent ρ ($U\varepsilon RV_\rho$) if

$$\lim_{t\to\infty} \frac{U(xt)}{U(t)} = x^\rho \text{ for all } x>0.$$

If U is monotone nonincreasing (as in the case of distribution tails), the convergence is uniform on $[x_0,\infty)$, $x_0>0$. If $\rho=0$, U is said to be slowly varying. We will occasionally use the fact that if U is nondecreasing and $U\varepsilon RV_\rho$, then for any $\beta>0$ there exists $K>0$ such that

$$\frac{U(xt)}{U(t)} \leq K \max(1,x^\beta) \text{ for all } x>0,\ t\geq 1.$$

Suppose $X_1,X_2,X_3,\ldots$ are independent, each with distribution F and suppose there exists sequences $\{a_n\},\{b_n\}$ such that

$$\frac{1}{a_n} \sum_{j=1}^{n} (X_j - b_n) \Rightarrow F_1$$

where F_1 is a distribution function. When this holds, F_1 is called a stable law and F is in the domain of attraction of F_1. Stable laws can be characterized nicely (see, e.g., Feller (1971)), but we mention only the following characterization of domains of attraction. Let $\mu(t) = \int_{-t}^{t} u^2 F(du)$. Then F is in a domain of attraction of a stable law if and only if for some $\alpha \varepsilon (0,2]$, $\mu \varepsilon RV_{2-\alpha}$, and if (when $\alpha<2$, $X \sim F$)

$$\lim_{t\to\infty} \frac{P[X>t]}{P[|X|>t]} = p \varepsilon [0,1].$$

In this case we write $F \varepsilon \mathcal{D}(\alpha)$. When $\alpha<2$, the tails of F are regularly varying, e.g. $P[|X|>t] \varepsilon RV_{-\alpha}$.

We leave the definition of bivariate stable domain of attraction until later.

In Section 2, we present two lemmas from the literature giving conditions for which $P[|XY|>t]$ is regularly varying. These lemmas will be used in Section 3 for our domain of attraction results. In Section 3, we assume $F \varepsilon \mathcal{D}(\alpha_1)$ and then with appropriate conditions on G, we will show $H \varepsilon \mathcal{D}(\alpha_1)$.

Section 4 extends the results of Section 3 to the bivariate case. In fact, the results presented here are easily extended to several variables.

The results of Part III will be used in Part IV for limit theorems concerning M-estimators of regression.

2. Regular Variation of the Product Distribution

We assume $X \sim F$ and $Y \sim G$ are independent random variables. We also assume X and Y are almost surely positive. Let H be the distribution of XY. We wish to provide conditions for which $P[XY>t] \in RV_{-\alpha}$. In the following, $\bar{F}(t) = 1-F(t)$, $\bar{G}(t) = 1-G(t)$ and $\bar{H}(t) = 1-H(t)$. The first lemma is due to Breiman (1965). Although he states it more narrowly than we do here, the two versions are equivalent.

Lemma 2.1 Suppose $\bar{F} \in RV_{-\alpha}$ and $EY^{\beta} < \infty$ for some $\beta > \alpha$. Then $\bar{H} \in RV_{-\alpha}$ and

$$\lim_{t\to\infty} \frac{P[XY>t]}{P[X>t]} = \lim_{t\to\infty} \int_0^{\infty} \frac{\bar{F}(t/y)}{\bar{F}(t)} G(dy) = \int_0^{\infty} y^{\alpha} G(dy).$$

Proof. i) Let $U(t) = \frac{1}{\bar{F}(t)}$. By the comment in the introduction, there exists K such that $U(xt) \le K \max(1,x^{\beta}) U(t)$ for all $x>0$, $t \ge 1$. Thus

$$\frac{\bar{F}(t/y)}{\bar{F}(t)} = \frac{U((t/y)y)}{U(t/y)} \le K \max(1,y^{\beta}) \text{ for } y \le t.$$

For $y>t$,

$$\frac{\bar{F}(t/y)}{\bar{F}(t)} \le \frac{1}{\bar{F}(t)}$$

$$\le \frac{y^{\beta}}{t^{\beta}\bar{F}(t)} \quad .$$

For sufficiently large t, however, $t^{\beta}\bar{F}(t) \ge 1/K$.

We therefore have for large enough t,

$$\frac{\bar{F}(t/y)}{\bar{F}(t)} \le K \max(1,y^{\beta}) \quad \text{for all } y.$$

By dominated convergence,

$$\lim_{t\to\infty} \int_0^{\infty} \frac{\bar{F}(t/y)}{\bar{F}(t)}\, G(dy) = \int_0^{\infty} \lim_{t\to\infty} \frac{\bar{F}(t/y)}{\bar{F}(t)}\, G(dy)$$

$$= \int_0^{\infty} y^{\alpha} G(dy). \qquad \#$$

That $EY^{\alpha}<\infty$ is not sufficient in Lemma 2.1 is illustrated in the following example. Suppose,

$$F(t) = \frac{e}{t(\ln t)^2}\, 1_{[e,\infty)}(t),$$

$$G(dt) = \frac{k}{(t \ln t)^2}\, 1_{[e,\infty)}(t)\, dt.$$

We note that

$$k = \left(\int_1^{\infty} \frac{1}{u^2}\, e^{-u}\, du\right)^{-1} = \int_0^{\infty} tG(dt)$$

and $\qquad e = \int_0^{\infty} t\, F(dt).$

Then, substituting $x = \frac{1}{\ln y}$,

$$\bar{H}(t) = \int_0^\infty \bar{F}(t/y)G(dy) = \int_e^{t/e} \frac{ey}{t(\ln t-\ln y)^2} \frac{k}{(y\ln y)^2}\,dy$$

$$= \frac{ek}{t} \int_{\frac{1}{\ln t-1}}^{1} \left(\frac{x}{x\ln t-1}\right)^2 dx$$

$$= \frac{2ek}{t(\ln t)^3}\left[\ln t-1+2\ln(\ln t-1) - \frac{1}{\ln t-1}\right]$$

$$\sim \frac{2ek}{t(\ln t)^2}$$

Therefore

$$\lim_{t\to\infty} \frac{\bar{H}(t)}{\bar{F}(t)} = 2\int_0^\infty t\,G(dt)$$

and

$$\lim_{t\to\infty} \frac{\bar{H}(t)}{\bar{G}(t)} = 2\int_0^\infty t\,F(dt).$$

Another lemma, due to Embrechts and Goldie (1980), gives the regular variation of $\bar{H}$, but we do not get the additional conclusion that $\bar{H}(t)/\bar{F}(t)$ has a limit. The proof of part ii) has been improved slightly.

Lemma 2.2 Suppose $\bar{F}\varepsilon RV_{-\alpha}$. Then $\bar{H}\varepsilon RV_{-\alpha}$ if either

i) $\lim_{t\to\infty} \frac{\bar{G}(t)}{\bar{F}(t)} = 0$

or

ii) $G\varepsilon RV_{-\alpha}$.

Proof. i) Choose s_0 such that $\bar{G}(1/s)>0$ for all $s \geq s_0$. Since

$\lim_{t\to\infty} \frac{P[Y>t]}{P[X>t]} = 0$, then

$$\overline{\lim_{t\to\infty}} \frac{P[XY>t,\ Y>t/s]}{P[XY>t,\ Y<t,s]} \leq \overline{\lim_{t\to\infty}} \frac{P[Y>t/s]}{P[X>ts]P[1/s<Y<t/s]}$$

$$= \overline{\lim_{t\to\infty}} \frac{P[Y>ts]}{P[X>ts]} \lim_{t\to\infty} \frac{P[Y>t/s]}{P[Y>ts]P[1/s<Y<t/s]}$$

$$= 0(s^{2\alpha}).$$

Therefore,

$$\lim_{t\to\infty} \int_0^{t/s} \frac{\bar{F}(t/y)}{\bar{H}(t)} G(dy) = \lim_{t\to\infty} \frac{P[XY>t,\ Y<t/s]}{P[XY>t]} = 1 \tag{2.1}$$

Replacing t with xt, s with xs, $x>0$, $s>s_0/x$, then

$$\lim_{t\to\infty} \int_0^{t/s} \frac{\bar{F}(xt/y)}{\bar{H}(xt)} G(dy) = 1. \tag{2.2}$$

Now let $m_F(s) = \inf_{u>s} \frac{\bar{F}(xu)}{\bar{F}(u)}$ and $M_F(x) = \sup_{u>s} \frac{\bar{F}(xu)}{\bar{F}(u)}$. Then

$\lim_{s\to\infty} m_F(s) = \lim_{s\to\infty} M_F(s) = x^{-\alpha}$ and for any $y<t/s$

$$m_F(s) \leq \frac{\bar{F}(xt/y)}{\bar{F}(t/y)} \leq M_F(s).$$

Therefore (2.1) and (2.2) give

$$m_F(s) \le \underline{\lim_{t\to\infty}} \frac{\int_0^{t/s} \bar{F}(xt/y)G(dy)}{\int_0^{t/s} \bar{F}(t/y)G(dy)}$$

$$= \underline{\lim_{t\to\infty}} \frac{\bar{H}(xt)}{\bar{H}(t)}$$

$$\le \overline{\lim_{t\to\infty}} \frac{\bar{H}(xt)}{\bar{H}(t)}$$

$$= \overline{\lim_{t\to\infty}} \frac{\int_0^{t/s} \bar{F}(xt/y)G(dy)}{\int_0^{t/s} \bar{F}(t/y)G(dy)}$$

$$\le M_F(s).$$

And this implies $\lim_{t\to\infty} \frac{\bar{H}(xt)}{\bar{H}(t)} = x^{-\alpha}$.

ii) Define m_G and M_G for G, just as m_F and M_F were defined above for F. For any $s>0$,

$$\bar{H}(t) = P[XY>t,\ Y\le t/s] + P[XY>t,\ X\le s] + P[X>s,\ Y>t/s]$$

$$= \int_0^{t/s} \bar{F}(t/y)G(dy) + \int_0^{s} \bar{G}(t/y)F(dy) + \bar{F}(s)\bar{G}(t/s). \qquad (2.3)$$

Replacing t with xt, s with xs, $0<x\le 1$,

$$\bar{H}(xt) = \int_0^{t/s} \bar{F}(xt/y)G(dy) + \int_0^{xs} \bar{G}(xt/y)F(dy) + \bar{F}(xs)\bar{G}(t/s)$$

$$\leq M_F(s)\int_0^{t/s} \bar{F}(t/y)G(dy) + M_G(t/xs)\int_0^{s} \bar{G}(t/y)F(dy)$$

$$+ M_F(s)\bar{F}(s)\bar{G}(t/s)$$

$$\leq \max(M_F(s), M_G(t/xs))\bar{H}(t) \tag{2.4}$$

Using (2.3) and (2.4)

$$\overline{\lim_{t\to\infty}} \frac{\bar{H}(xt)}{\bar{H}(t)} \leq \lim_{t\to\infty} \max(M_F(x),\ M_G(t/xs))$$

$$= \max(M_F(s), x^{-\alpha})$$

and therefore, letting $s\to\infty$,

$$\lim_{t\to\infty} \frac{\bar{H}(xt)}{H(t)} \leq x^{-\alpha} \text{ for } x\varepsilon(0,1]. \tag{2.5}$$

Similarly,

$$\bar{H}(xt) \geq \min(m_F(s), m_G(t,xs))\bar{H}(t) - \int_{xs}^{s} \bar{G}(xt/y)F(dy)$$

and since s is arbitrary

$$\lim_{t\to\infty} \frac{\bar{H}(xt)}{\bar{H}(t)} \geq \overline{\lim_{s\to\infty}} \lim_{t\to\infty} \left[\min(m_F(s), m_G(t/xs)) - \int_{xs}^{s} \frac{\bar{G}(xt/y)}{\bar{H}(t)} F(dy)\right]$$

$$= x^{-\alpha} - \overline{\lim_{s\to\infty}} \lim_{t\to\infty} \int_{xs}^{s} \frac{\bar{G}(xt/y)}{\bar{H}(t)} F(dy) \tag{2.6}$$

To handle the second term, we note that

$$\int_{xs}^{s} \bar{G}(xt/y)F(dy) = P[XY>xt,\ xs<X<s]$$

$$\leq P[X>xs]P[Y>xt/s]$$

$$= \bar{F}(xs)\bar{G}(xt/s). \tag{2.7}$$

However, by Fatou's Lemma,

$$\varliminf_{t\to\infty} \frac{\bar{H}(t)}{\bar{G}(xt/s)} \geq \int_0^\infty \varliminf_{t\to\infty} \frac{\bar{G}(t/y)}{\bar{G}(xt/s)} F(dy)$$

$$= \int_0^\infty \left(\frac{xy}{s}\right)^\alpha F(dy)$$

$$= \frac{x^\alpha EX^\alpha}{s^\alpha}, \tag{2.8}$$

which may be infinite. If $EX^\alpha<\infty$, $\lim\limits_{s\to\infty} s^\alpha \bar{F}(xs) = 0$. In either case

$$\lim_{s\to\infty} \varlimsup_{t\to\infty} \frac{\bar{G}(xt/s)\bar{F}(xs)}{\bar{H}(t)} \leq \lim_{s\to\infty} \frac{s^\alpha \bar{F}(xs)}{x^\alpha EX^\alpha}$$

$$= 0\ . \tag{2.9}$$

Using (2.7) and (2.9) in (2.6)

$$\varliminf_{t\to\infty} \frac{\bar{H}(xt)}{\bar{H}(t)} \geq x^{-\alpha}$$

and with (2.5)

$$\lim_{t\to\infty} \frac{\bar{H}(xt)}{\bar{H}(t)} = x^{-\alpha} \text{ for } x\varepsilon(0,1].$$

The case x>1 is a simple consequence of this. #

3. Stable Attraction of the Product Distribution

Continuing the nomenclature in Section 2, we now wish to determine conditions for which $H\varepsilon\mathcal{D}(\alpha)$ for some $\alpha\leq 2$. The cases $\alpha<2$ rely on Lemmas 2.1 and 2.2 and the case $\alpha=2$ is handled in a manner similar to the proofs of those lemmas. The random variables may now be negative or positive.

Theorem 3.1 Suppose $F\varepsilon\mathcal{D}(a)$, $\alpha<2$ and $E|Y|^{\beta}<\infty$ for some $\beta>\alpha$. Then $H\varepsilon\mathcal{D}(\alpha)$.

Proof. Define F^* and H^* to be the distributions of $|X|$ and $|XY|$, respectively. By Lemma 2.1 we have

$$\lim_{t\to\infty} \frac{P[|XY|>t]}{P[|X|>t]} = \lim_{t\to\infty} \frac{\bar{H}^*(t)}{\bar{F}^*(t)} = E|Y|^{\alpha}. \tag{3.1}$$

Since $F\varepsilon\mathcal{D}(\alpha)$, $\alpha<2$ then $\bar{F}^*\varepsilon RV_{-\alpha}$ and hence by (3.1)

$$\lim_{t\to\infty} \frac{\bar{H}^*(xt)}{\bar{H}^*(t)} = \lim_{t\to\infty} \frac{\bar{F}^*(xt)}{\bar{F}^*(t)} = x^{-\alpha}.$$

That is, $\bar{H}^*$ is regularly varying with exponent $-\alpha$. To show that H is in a domain of attraction, it remains only to show that $\lim_{t\to\infty} \frac{\bar{H}(t)}{\bar{H}^*(t)}$ exists.

Let $p = \lim_{t\to\infty} \frac{\bar{F}(t)}{\bar{F}^*(t)}$, which exists since $F\varepsilon\mathcal{D}(\alpha)$, $\alpha<2$. Define also

$$q = \frac{pE[Y^{\alpha}1_{Y\geq 0}] + (1-p)E[(-Y)^{\alpha}1_{Y<0}]}{E|Y|^{\alpha}}. \tag{3.2}$$

A dominated convergence argument like that in the proof of Lemma 2.1 will work, but we present another argument here. By Fatou's Lemma,

$$\varliminf_{t\to\infty} \frac{P[XY>t, Y\geq 0]}{P[|XY|>t]} = \varliminf_{t\to\infty} \int_0^\infty \frac{\bar{F}(t/y)}{\bar{H}^*(t)} G(dy)$$

$$\geq \int_0^\infty \varliminf_{t\to\infty} \frac{\bar{F}(t/y)}{\bar{H}^*(t)} G(dy)$$

$$= \frac{p\ E[Y^\alpha 1_{Y\geq 0}]}{E|Y|^\alpha} \tag{3.3a}$$

Similarly, we have

$$\varliminf_{t\to\infty} \frac{P[XY>t, Y<0]}{P[|XY|>t]} \geq \frac{(1-p)E[(-Y)^\alpha 1_{Y<0}]}{E|Y|^\alpha}, \tag{3.3b}$$

$$\varliminf_{t\to\infty} \frac{P[XY<-t, Y\geq 0]}{P[|XY|>t]} \geq \frac{(1-p)E[Y^\alpha 1_{Y\geq 0}]}{E|Y|^\alpha} \tag{3.3c}$$

and

$$\varliminf_{t\to\infty} \frac{P[XY<-t, Y<0]}{P[|XY|>t]} \geq \frac{p\ E[(-Y)^\alpha 1_{Y<0}]}{E|Y|^\alpha}. \tag{3.3d}$$

By adding (3.3a) and (3.3b), we obtain

$$\varliminf_{t\to\infty} \frac{\bar{H}(t)}{\bar{H}^*(t)} \geq q \tag{3.4}$$

where q is defined in (3.2). By adding (3.3c) and (3.3d) we have

$$\varliminf_{t\to\infty} \left(1 - \frac{\bar{H}(t)}{\bar{H}^*(t)}\right) \geq 1-q. \tag{3.5}$$

Together (3.4) and (3.5) imply

$$\lim_{t\to\infty} \frac{\bar{H}(t)}{\bar{H}^*(t)} = q.$$

Hence $H\varepsilon\mathcal{D}(\alpha)$. #

Corollary 3.2 Suppose $F\varepsilon\mathcal{D}(\alpha)$, $G\varepsilon\mathcal{D}(\beta)$, $\beta>\alpha$. Then $H\varepsilon\mathcal{D}(\alpha)$.

Proof. Use Theorem 3.1 and the fact that $E|Y|^{(\beta+\alpha)/2}<\infty$. #

When F and G are in domains of attraction with the same index α, we have

Theorem 3.3 Suppose $F\varepsilon\mathcal{D}(\alpha)$, $G\varepsilon\mathcal{D}(\alpha)$. Then $H\varepsilon\mathcal{D}(\alpha)$ provided one of the following is true

i) either F or G is symmetric,

ii) both F and G concentrate on $[0,\infty)$,

iii) $\lim_{t\to\infty} \frac{P[|XY|>t]}{P[|X|>t]} = E|Y|^{\alpha}<\infty$, $\alpha<2$,

iv) $E|X|^{\alpha} = E|Y|^{\alpha} = \infty$, $\alpha<2$,

or v) $\alpha=2$.

Proof. Define F^*, G^* and H^* to be the distributions of $|X|$, $|Y|$ and $|XY|$, respectively.

Suppose $\alpha<2$. Then by Lemma 2.2 $\bar{H}^*\varepsilon RV_{-\alpha}$, since $\bar{F}^*\varepsilon RV_{-\alpha}$ and $\bar{G}^*\varepsilon RV_{-\alpha}$.

To show $H\varepsilon\mathcal{D}(\alpha)$, therefore, we need only to show that $\lim_{t\to\infty} \frac{\bar{H}(t)}{\bar{H}^*(t)}$ exists.

If either F or G is symmetric (case i), then certainly H is also, so that

$$\lim_{t\to\infty} \frac{\bar{H}(t)}{\bar{H}^*(t)} = \frac{1}{2} .$$

If both F and G concentrate on $[0,\infty)$ (case ii), then clearly

$$\lim_{t\to\infty} \frac{\bar{H}(t)}{\bar{H}^*(t)} = 1.$$

Case iii) If $\lim_{t\to\infty} \frac{P[|XY|>t]}{P[|X|>t]} = E|Y|^{\alpha}<\infty$, $\alpha<2$, then the proof of Theorem 3.1 is still valid. Hence,

$$\lim_{t\to\infty} \frac{\bar{H}(t)}{\bar{H}^*(t)} = q$$

where q is defined in (3.2).

iv) Suppose $E|X|^{\alpha} = E|Y|^{\alpha} = \infty$, $\alpha<2$. Let

$$p_1 = \lim_{t\to\infty} \frac{\bar{F}(t)}{\bar{F}^*(t)}$$

and

$$p_2 = \lim_{t\to\infty} \frac{\bar{G}(t)}{\bar{G}^*(t)} \quad .$$

Now choose $t_0>0$ such that for $t>t_0$ $(0<\varepsilon<\frac{1}{2})$

$$(1-\varepsilon)p_1\bar{F}^*(t) \leq \bar{F}(t) \leq (1+\varepsilon)p_1\bar{F}^*(t)$$

and

$$(1-\varepsilon)p_2\bar{G}^*(t) \leq \bar{G}(t) \leq (1+\varepsilon)p_2\bar{G}^*(t).$$

Then for any $s>t_0$, $t>s^2$,

$$\begin{aligned}\int_s^{t/s} \bar{G}(t/u)F(du) &= P[X>s,Y>t/s] - P[X>t/s,Y>s] + P[XY>t,s<Y<t/s] \\ &= \bar{F}(s)\bar{G}(t/s) - \bar{F}(t/s)\bar{G}(s) + \int_s^{t/s} \bar{F}(t/u)G(du)\end{aligned}$$

$$\leq (1+\varepsilon)p_1\left[\bar{F}^*(s)\bar{G}(t/s) + \int_s^{t/s} \bar{F}^*(t/u)G(du)\right] - (1-\varepsilon)p_1\bar{F}^*(t/s)\bar{G}(s)$$

$$= (1+\varepsilon)p_1\int_s^{t/s} \bar{G}(t/u)F^*(du) + 2\varepsilon p_1\bar{F}^*(t/s)\bar{G}(s)$$

$$\leq (1+\varepsilon)^2 p_1 p_2 \int_s^{t/s} \bar{G}^*(t/u)F^*(du) + (1+\varepsilon)\bar{F}^*(t/s).$$

This implies

$$P[XY>t, X\geq 0] = \int_0^s \bar{G}(t/u)F(du) + \int_0^s \bar{F}(t/u)G(du) + \int_s^{t/s} \bar{G}(t/u)F(du)$$

$$+ \bar{F}(t/s)\bar{G}(s)$$

$$\leq (1+\varepsilon)[p_2\bar{G}^*(t/s) + p_1\bar{F}^*(t/s)]$$

$$+ (1+\varepsilon)^2 p_1 p_2 \left[\int_s^{t/2} \bar{G}^*(t/u)F^*(du) + \bar{F}^*(t/s)\bar{G}^*(s)\right] + \bar{F}^*(t/s)$$

$$\leq (1+\varepsilon)[\bar{G}^*(t/s) + 2\bar{F}^*(t/s)] + (1+\varepsilon)^2 p_1 p_2 \; P[|XY|>t].$$

And therefore

$$\overline{\lim_{t\to\infty}} \frac{P[XY>t, X>0]}{P[|XY|>t]} \leq (1+\varepsilon)^2 p_1 p_2 + (1+\varepsilon) \overline{\lim_{t\to\infty}} \frac{\bar{G}^*(t/s) + 2\bar{F}^*(t/s)}{\bar{H}^*(t)} . \tag{3.6}$$

We recall (2.8) which, slightly revised, says

$$\overline{\lim_{t\to\infty}} \frac{\bar{G}^*(t/s)}{\bar{H}^*(t)} \leq \frac{s^\alpha}{E|X|^\alpha} = 0.$$

Similarly

$$\overline{\lim_{t\to\infty}} \frac{\bar{F}^*(t/s)}{\bar{H}^*(t)} \le \frac{s^\alpha}{E|Y|^\alpha} = 0.$$

Since ε is arbitrary, (3.6) becomes therefore

$$\overline{\lim_{t\to\infty}} \frac{P[XY>t, X\geq 0]}{P[|XY|>t]} \le p_1 p_2. \tag{3.7}$$

In a similar manner we can show for t large enough,

$$P[XY>t, X>0] \ge (1-\varepsilon)^2 p_1 p_2 \; P[|XY|>t] - (1+\varepsilon)(2\bar{F}^*(t/s) + \bar{G}^*(t/s))$$

so that

$$\underline{\lim_{t\to\infty}} \frac{P[XY>t, X\geq 0]}{P[|XY|>t]} \ge p_1 p_2.$$

With (3.7), this yields

$$\lim_{t\to\infty} \frac{P[XY>t, X\geq 0]}{P[|XY|>t]} = p_1 p_2. \tag{3.8}$$

By an identical argument,

$$\lim_{t\to\infty} \frac{P[XY>t, X<0]}{P[|XY|>t]} = (1-p_1)(1-p_2)$$

and therefore

$$\lim_{t\to\infty} \frac{\bar{H}(t)}{\bar{H}^*(t)} = \lim_{t\to\infty} \frac{P[XY>t]}{P[|XY|>t]} = p_1 p_2 + (1-p_1)(1-p_2).$$

The fact that the limit exists is all that we yet needed to show $H \varepsilon \mathcal{D}(\alpha)$. Note that this is not the same limit obtained in case iii). This completes the argument for the cases where $\alpha<2$.

v) Suppose finally that $\alpha=2$. Let $\mu_1(t) = \int_{-t}^{t} u^2F(du)$ and $\mu_2(t) = \int_{-t}^{t} u^2G(du)$. Also let F* and G* be as above. We are assuming that μ_1 and μ_2 are slowly varying and we want to show that $\mu(t) = \int_{-t}^{t} u^2H(du)$ is also. First note that for any $s>0$

$$\mu(t) = E[X^2 1_{|X|\le t/s}]E[Y^2 1_{|Y|\le s}] + E[(XY)^2 1_{|X|\le t/|Y|, |Y|>s}]$$

$$+ E[(XY)^2 1_{|Y|\le t/|X|, |X|>t/s}]$$

$$\mu_1(t/s)\mu_2(s) + \int_s^\infty \mu_1(t/u)u^2G^*(du) + \int_{t/s}^\infty \mu_2(t/u)u^2F^*(du). \tag{3.9}$$

Replacing t with xt, s with xs, $x\ge 1$,

$$\mu(xt) = \mu_1(t/s)\ \mu_2(xs) + \int_{xs}^\infty \mu_1(xt/u)u^2G^*(du)$$

$$+ \int_{t/s}^\infty \mu_2(xt/u)u^2F^*(du). \tag{3.10}$$

Define

$$m_i(s) = \inf_{u\ge s} \frac{\mu_i(xu)}{\mu_i(u)}$$

$$M_i(s) = \sup_{u\ge s} \frac{\mu_i(xu)}{\mu_i(u)}$$

Then by (3.9) and (3.10)

$$\mu(xt) \le M_2(s)\left[\mu_1(t/s)\mu_2(s) + \int_{t/s}^{\infty} \mu_2(t/u)u^2F^*(du)\right]$$

$$+ M_1(s) \int_s^{\infty} \mu_1(t/u)u^2G^*(du)$$

$$\le \max(M_1(s),M_2(s))\mu(t).$$

Since s is arbitrary, then

$$\overline{\lim_{t\to\infty}} \frac{\mu(xt)}{\mu(t)} \le \lim_{s\to\infty} \max(M_1(s),M_2(s))$$

$$= 1, \ x\ge 1. \tag{3.11}$$

On the other hand,

$$\mu(xt) \ge m_2(s)\left[\mu_1(t/s)\mu_2(s) + \int_{t/s}^{\infty} \mu_2(t/u)u^2F^*(du)\right]$$

$$+ m_1(s) \int_{xs}^{\infty} \mu_1(t/u)u^2G^*(du)$$

$$\ge \min(m_1(s),m_2(s))\mu(t) - m_1(s) \int_s^{xs} \mu_1(t/u)u^2G^*(du).$$

And thus, since s is arbitrary,

$$\underline{\lim_{t\to\infty}} \frac{\mu(xt)}{\mu(t)} \ge \overline{\lim_{s\to\infty}} \left[\min(m_1(s),m_2(s) - m_1(s) \overline{\lim_{t\to\infty}} \int_s^{xs} \frac{\mu_1(t/u)}{\mu(t)} u^2G^*(du)\right]$$

$$= 1 - \underline{\lim_{s\to\infty}} \overline{\lim_{t\to\infty}} \int_s^{xs} \frac{\mu_1(t/u)}{\mu(t)} u^2G(du).$$

Since μ_1 is nondecreasing,

$$\underline{\lim_{t\to\infty}} \frac{\mu(xt)}{\mu(t)} \ge 1 - \underline{\lim_{s\to\infty}} \overline{\lim_{t\to\infty}} \frac{\mu_1(t/s)}{\mu(t)} \int_s^{xs} u^2G(du). \tag{3.12}$$

By Fatou's Lemma, however,

$$\varliminf_{t\to\infty} \frac{\mu(t)}{\mu_1(t)} \geq \int_0^\infty \varliminf_{t\to\infty} \frac{\mu_1(t/u)}{\mu_1(t)}\, u^2 G(du)$$

$$= EY^2,$$

which may be infinite. If $EY^2<\infty$, then $\lim\limits_{s\to\infty} \int_s^{xs} u^2 G(du) = 0$. In either case, we have

$$\varliminf_{s\to\infty} \left[\varlimsup_{t\to\infty} \frac{\mu_1(t/s)}{\mu(t)}\right] \int_s^{xs} u^2 G(du) \leq \varliminf_{s\to\infty} \frac{1}{EY^2} \int_s^{xs} u^2 G(du) = 0$$

Thus (3.12) becomes

$$\varliminf_{t\to\infty} \frac{\mu(xt)}{\mu(t)} \geq 1.$$

With (3.11), we have

$$\lim_{t\to\infty} \frac{\mu(xt)}{\mu(t)} = 1, \; x\geq 1.$$

And this holds, in fact, for all $x>0$ as a simple consequence. Therefore, μ is slowly varying and $H \varepsilon \mathcal{D}(2)$. #

4. Joint Attraction of Two Products

Suppose now that we have $(X_1,X_2)\sim F$ and $(Y_1,Y_2)\sim G$ are independent. Let H be the distribution of (X_1Y_1,X_2Y_2). We shall use F_i,G_i,H_i, $i=1,2$ to denote the marginal distributions. In Section 3 we presented conditions for which each marginal distribution H_i is in a domain of attraction. Here we wish to discuss domains of attraction for bivariate distributions, and give conditions on F and G for which H is in a bivariate domain of attraction.

We say bivariate distribution F is in a bivariate domain of attraction if for a sequence $\{(X_{1n},X_{2n})\}$ of independent pairs of random variables, distributed by F, there exists sequences $\{a_{in},b_{in}\}$, $i=1,2$, such that the normalized partial sums

$$\frac{1}{a_{1n}}\sum_{j=1}^{n}(X_{1j}-b_{1n}),\ \frac{1}{a_{2n}}\sum_{j=1}^{n}(X_{2j}-b_{2n})$$

converge jointly to some bivariate distribution which has nondegenerate marginal distributions. (Clearly, the marginals are each univariate stable.) The class of such limit distributions is called the class of bivariate stable distributions. We shall write $F\varepsilon\mathcal{D}(\alpha_1,\alpha_2)$ to indicate that F is in the domain of attraction of a bivariate stable whose marginals are stable (α_1) and stable (α_2). The condition $F\varepsilon\mathcal{D}(\alpha_1,\alpha_2)$ is equivalent to one of the following, depending on the values of α_1 and α_2 (c.f. Resnick and Greenwood (1979) and deHaan, Omey and Resnick (1982)). Let $(X_1,X_2)\sim F$.

i) $\alpha_1<2,\alpha_2<2$. $F\varepsilon\mathcal{D}(\alpha_1,\alpha_2)$ if and only if

$$\lim_{n\to\infty} nP\left[\left(\frac{X_1}{a_{1n}},\frac{X_2}{a_{2n}}\right)\varepsilon A\right]=\nu(A) \tag{4.1}$$

for some sequence $\{a_{1n},a_{2n}\}$ and every Borel set $A\varepsilon\mathcal{R}^2-\{0\}$ such that the boundary of A has ν-measure zero. The limiting measure ν is called the Levv measure of the limiting stable distribution.

ii) $\alpha_1<\alpha_2=2$. $F\varepsilon\mathcal{D}(\alpha_1,2)$ if and only if $F_1\varepsilon\mathcal{D}(\alpha_1)$ and $F_2\varepsilon\mathcal{D}(2)$. The limiting distribution is the joint distribution of independent (nonnormal) stable and normal random variables.

iii) $\alpha_1=\alpha_2=2$. $F\varepsilon\mathcal{D}(2,2)$ if and only if each marginal F_1 and F_2 is in $\mathcal{D}(2)$ and for the corresponding sequences $\{a_{in}\}$, i=1,2, (assume $EX_1=EX_2=0$)

$$\lim_{n\to\infty}\frac{nE\left[X_1X_2\,1_{|X_1|\le a_{1n}u_1}\,1_{|X_2|\le a_{2n}u_2}\right]}{a_{1n}\,a_{2n}}=c_{12} \tag{4.2}$$

for any $u_1>0,u_2>0$. In this case c_{12} is the asymptotic covariance and the asymptotic variances are

$$c_{ii}=\lim_{n\to\infty}\frac{nE\left[X_i^2 1_{|X_i|\le a_{in}}\right]}{a_{in}^2}\ .$$

To facilitate our use of condition i) we give the following further characterization.

<u>Lemma 4.1</u> Suppose $F_i\varepsilon\mathcal{D}(\alpha_i),\alpha_i<2$, and define for $X\varepsilon\mathcal{R}$

$$U_i(x)=\frac{\operatorname{sgn}(x)}{P[|X_i|>|x|]}\ ,\ i=1,2\ .$$

Then $F\varepsilon\mathcal{D}(\alpha_1,\alpha_2)$ if and only if for every $c_1,c_2\neq 0$,

$$\lim_{n\to\infty} nP[\max(c_1U_1(X_1),c_2U_2(X_2))>n] = \Pi(c_1,c_2)$$

for some (necessarily continuous) function Π. In this case the convergence is uniform for $|c_i|\geq c_0>0$ and

$$\Pi(c_1,c_2) = \nu(\{(x_1,x_2):\max(c_1[x_1]^{1/\alpha_1},c_2[x_2]^{1/\alpha_2}) > 1\}),$$

where $[x]^\gamma = \mathrm{sgn}(x)|x|^\gamma$ for any $\gamma>0$.

<u>Proof</u>. We follow a similar argument outlined by Resnick and Greenwood (1979). Assume first that $F\varepsilon\mathcal{D}(\alpha_1,\alpha_2)$. Then there exist sequences $\{a_{in}\}$ such that (4.1) is true. In fact since each marginal F_i is in $\mathcal{D}(\alpha_i)$, we can choose a_{in} to satisfy

$$\lim_{n\to\infty} nP\left[\left|\frac{X_i}{a_{in}}\right|>1\right] = 1.$$

By the regular variation property of the tails of F_i,

$$\lim_{n\to\infty} nP\left[\left|\frac{X_i}{a_{in}}\right|>x\right] = x^{-\alpha_i},\ x>0. \tag{4.3}$$

This implies, for $0<\varepsilon<1/2$

$$\begin{aligned}\lim_{n\to\infty} nP\left[1-\varepsilon<\left|\frac{c_iX_i}{a_{in}}\right|<1+\varepsilon\right] &= |c_i|^{\alpha_i}\left[(1-\varepsilon)^{-\alpha_i}-(1+\varepsilon)^{-\alpha_i}\right]\\ &\leq 4\alpha_i|c_i|^{\alpha_i}\varepsilon.\end{aligned} \tag{4.4}$$

Equation (4.4) tells us two things. First,

$$\lim_{n\to\infty} nP\left[1-\varepsilon < \max\left(\frac{c_1X_1}{a_{1n}}, \frac{c_2X_2}{a_{2n}}\right) < 1+\varepsilon\right]$$

$$\leq \lim_{n\to\infty} nP\left[1-\varepsilon < \left|\frac{c_1X_1}{a_{1n}}\right| < 1+\varepsilon\right] + \lim_{n\to\infty} nP\left[1-\varepsilon < \left|\frac{c_2X_2}{a_{2n}}\right| < 1+\varepsilon\right]$$

$$\leq 4(\alpha_1|c_1|^{\alpha_1} + \alpha_2|c_2|^{\alpha_2})\varepsilon$$

and therefore

$$\lim_{n\to\infty} nP\left[\max\left(\frac{c_1X_1}{a_{1n}}, \frac{c_2X_2}{a_{2n}}\right) = 1\right] = 0.$$

Letting $A(c_1,c_2) = \{(x_1,x_2):\max(c_1x_1,c_2x_2) > 1\}$, we can now apply (4.1) and obtain

$$\lim_{n\to\infty} nP\left[\max\left(\frac{c_1X_1}{a_{1n}}, \frac{c_2X_2}{a_{2n}}\right) > 1\right] = \nu(A(c_1,c_2)). \tag{4.5}$$

Second, (4.4) tells us that the limit in (4.5), $(A(c_1,c_2))$ is a continuous function in c_1 and c_2. Because both the sequence and the limit are monotone functions for $c_i>0$ as well, then the convergence is locally uniform. This generalizes, in fact, to the statement that (4.5) holds uniformly when $|c_i|\geq x_0$ for any $x_0>0$, since the functions are decreasing to zero as $|c_i|\to\infty$.

The definition of U_i and (4.3) give

$$\lim_{n\to\infty} \frac{1}{n} U_i(a_{in}x) = [x]^{\alpha_i}. \tag{4.6}$$

Taking inverses (valid by deHaan (1970), p.22 and the antisymmetry of U_i),

$$\lim_{n\to\infty} \frac{U_i^{\leftarrow}(nx)}{a_{in}} = [x]^{1/\alpha_i}. \tag{4.7}$$

Using (4.6), the uniform convergence in (4.5) and the monotonicity of U_i we obtain

$$\lim_{n\to\infty} nP[\max(c_1U_1(X_1), c_2U_2(X_2)) > n]$$

$$= \lim_{n\to\infty} nP\left[\max\left(\frac{X_1}{U_1^{\leftarrow}(n/c_1)}, \frac{X_2}{U_2^{\leftarrow}(n/c_2)}\right) > 1\right]$$

$$= \lim_{n\to\infty} nP\left[\max\left(\frac{[c_1]^{1/\alpha_1}X_1}{a_{1n}}, \frac{[c_2]^{1/\alpha_2}X_2}{a_{2n}}\right) > 1\right]$$

$$= \nu(A([c_1]^{1/\alpha_1}, [c_2]^{1/\alpha_2}))$$

$$= \nu(\{(x_1,x_2): \max(c_1[x_1]^{\alpha_1}, c_2[x_2]^{\alpha_2}) > 1\}).$$

Furthermore, the convergence is uniform for $|c_i| \geq x_0$ and the limit is continuous. This proves the first half of the lemma.

Now suppose $F_i \varepsilon \mathcal{D}(\alpha_i)$, $\alpha_i < 2$ and

$$\lim_{n\to\infty} nP[\max(c_1U_1(X_1), c_2U_2(X_2)) > n] = \Pi(c_1, c_2) \tag{4.8}$$

for some function Π and $c_i \neq 0$, $i=1,2$. This yields immediately that $\Pi(c_1k, c_2k) = k\Pi(c_1, c_2)$ for any $k>0$, and hence that Π is continuous. Since $\Pi \downarrow 0$ as $|c_i| \uparrow \infty$, then we have that the convergence in (4.8) is uniform for $|c_i| \geq x_0, x_0 > 0$. Again (4.6) holds because each marginal F_i is in $\mathcal{D}(\alpha_i)$, so that by the uniform convergence

$$\lim_{n\to\infty} nP\left[\max\left(\frac{c_1X_1}{a_{1n}},\frac{c_2X_2}{a_{2n}}\right)>1\right]$$

$$= \lim_{n\to\infty} nP\left[\max\left(\frac{U_1(X_1)}{U_1(a_{1n}/c_1)},\frac{U_2(X_2)}{U_2(a_{2n}/c_2)}\right)>1\right]$$

$$= \lim_{n\to\infty} nP\left[\max\left(\frac{[c_1]^{\alpha_1}U_1(X_1)}{n},\frac{[c_2]^{\alpha_2}U_2(X_2)}{n}\right)>1\right]$$

$$= \Pi([c_1]^{\alpha_1},[c_2]^{\alpha_2}). \tag{4.9}$$

For $A(c_1,c_2) = \{(x_1,x_2):\max(c_1x_1,c_2x_2)>1\}$. Define $\nu(A(c_1,c_2)) = \Pi([c_1]^{\alpha_1},[c_2]^{\alpha_2})$. Since the class of sets $A(c_1,c_2)$, $c_i\neq 0$ form a Π-class generating the Borel sets in $\mathcal{R}^2-\{0\}$, the usual class arguments show that ν can be extended to be a measure on $\mathcal{R}^2-\{0\}$ and that

$$\lim_{n\to\infty} nP\left[\left(\frac{X_1}{a_{1n}},\frac{X_1}{a_{2n}}\right)\varepsilon\ A\right] = \nu(A)$$

for A such that the boundary of A has ν-measure 0. Hence, $F\varepsilon\mathcal{D}(\alpha_1,\alpha_2)$.

#

The conditions i) - iii), as well as the next theorem, can easily be extended to more variables in the obvious manner. If the limiting distribution has normal and nonnormal components, then it factors accordingly. (This was proved by Sharpe (1969).)

<u>Theorem 4.2</u> Suppose $(X_1,X_2)\sim F$ where F is in a bivariate domain of attraction ($F\varepsilon\mathcal{D}(\alpha_1,\alpha_2)$) and suppose (X_1,X_2) is independent of (Y_1,Y_2), Y_i not degenerate at 0 and $E|Y_i|^{\beta_i}<\infty$ for some $\beta_i>\alpha_i$ (or $\beta_i=2$ if $\alpha_i=2$).

Let H be the distribution of (X_1Y_1, X_2Y_2). Then $H \varepsilon \mathcal{D}(\alpha_1, \alpha_2)$.

Proof: We consider the cases i) - iii) outlined above.

i) $\alpha_1<2, \alpha_2<2$. If H_i is the marginal distribution of X_iY_i, then Theorem 3.1 tells us immediately that $H_i \varepsilon \mathcal{D}(\alpha_i)$. Defining U_1 and U_2 as in Lemma 4.1, we have

$$\lim_{n\to\infty} nP[\max(c_1U_1(X_1), c_2U_2(X_2))>n] = \Pi(c_1, c_2) \tag{4.10}$$

for c_i not both zero and some function Π. This holds locally uniformly. Define also

$$V_i(x) = \frac{\operatorname{sgn}(x)}{P[|X_iY_i|>|x|]}$$

we need to show

$$\lim_{n\to\infty} nP[\max(c_1V_1(X_1Y_1), c_2V_2(X_2Y_2))>n] = Q(c_1, c_2)$$

for some function Q.

First, we fix y_i, not both zero. Since $F_i \varepsilon \mathcal{D}(\alpha_i)$ then each U_i is regularly varying on $(0,\infty)$. The sequences $a_{in} = U_i^{\leftarrow}(n)$ are acceptable normalizations. By Lemma 2.1, we also have that V_i is regularly varying and

$$\lim_{n\to\infty} \frac{U_i(n)}{V_i(n)} = E|Y_i|^{\alpha_i} .$$

Taking inverses (deHaan (1970), p. 22)

$$\lim_{n\to\infty} \frac{U_i^{\leftarrow}(n)}{V_i^{\leftarrow}(n)} = (E|Y_i|^{\alpha_i})^{-1/\alpha_i} .$$

The inverse $\overleftarrow{V}_i$ is regularly varying with exponent $1/\alpha_i$ on $(0,\infty)$ (again, c.f. deHaan) and is antisymmetric about 0, so that for $c_i \neq 0$

$$\lim_{n\to\infty} \frac{\overleftarrow{V}_i(n)}{\overleftarrow{V}_i(n/c_i)} = [c_i]^{1/\alpha_i}$$

Define $m_i = y_i[c_i]^{1/\alpha_i}(E|Y_i|^{\alpha_i})^{-1/\alpha_i}$. Then

$$\lim_{n\to\infty} \frac{y_i \overleftarrow{U}_i(n)}{\overleftarrow{V}_i(n/c_i)} = m_i \ . \tag{4.11}$$

By virtue of (4.11) and the fact that (4.10) holds locally uniformly,

$$\begin{aligned}
&\lim_{n\to\infty} nP[\max(c_1V_1(X_1y_1), c_2V_2(X_2y_2)) > n] \\
&\quad = \lim_{n\to\infty} nP\left[\max\left(\frac{y_1X_1}{\overleftarrow{V}_1(n/c_1)}, \frac{y_2X_2}{\overleftarrow{V}_2(n/c_2)}\right) > 1\right] \\
&\quad = \lim_{n\to\infty} nP\left[\max\left(\frac{m_1X_1}{\overleftarrow{U}_1(n)}, \frac{m_2X_2}{\overleftarrow{U}_2(n)}\right) > 1\right].
\end{aligned}$$

Relying on (4.9) from Lemma 4.1 and the normalizations $\overleftarrow{U}_i(n)$,

$$\begin{aligned}
&\lim_{n\to\infty} nP[\max(c_1V_1(X_1y_1), c_2V_2(X_2y_2)) > n] \\
&\quad = \Pi([m_1]^{\alpha_1}, [m_2]^{\alpha_2}) \\
&\quad = \Pi\left(c_1 \frac{[y_1]^{\alpha_1}}{E|Y_1|^{\alpha_1}}, c_2 \frac{[y_2]^{\alpha_2}}{E|Y_2|^{\alpha_2}}\right)
\end{aligned} \tag{4.12}$$

Of course, the limit is zero for $y_1 = y_2 = 0$.

Define $P^*(n) = P[\max(V_1(|X_1|), V_2(|X_2|)) > n]$ and Π^* by

$$\Pi^*(c_1, c_2) = \Pi(c_1, c_2) + \Pi(-c_1, c_2) + \Pi(c_1, -c_2) + \Pi(-c_1, -c_2)$$

By (4.12)

$$\lim_{n\to\infty} nP^*(n) = \Pi^*\left(\frac{1}{E|Y_1|^{\alpha_1}}, \frac{1}{E|Y_2|^{\alpha_2}}\right),$$

which says that P^* is regularly varying with exponent -1. Since it is also monotone we make use of a comment made in the introduction and used similarly in Lemma 2.1. That is, choose $\varepsilon > 0$. Then there exists K_1 such that for all $y>0$, n large enough,

$$P^*(n/y) \leq K_1 \max(1, |y|^{1+\varepsilon}) P^*(n) . \tag{4.13}$$

We will actually choose ε so that $\gamma_i = \beta_i/(1+\varepsilon) > \alpha_i$, where β_i are given in the theorem statement. Similarly, since V_i is regularly varying with exponent α_i, then since $\gamma_i > \alpha_i$, there exists $K_2 \geq 1$ such that for any $y>0$, $x \geq 1$, fixed c_i,

$$|c_i| V_i(xy) \leq K_2 \max(1, y^{\gamma_i}) V_i(x).$$

If $x<1$, then $|c_i| V_i(xy) \leq K_2 \max(1, y^{\gamma_i}) V_i(1)$, since V_i is monotone. Therefore using these inequalities and (4.13)

$$P[\max(c_1 V_1(X_1 y_1), c_2 V_2(X_2 y_2)) > n]$$

$$\leq P[\max(|c_1| V_1(|X_1 y_1|), |c_2| V_2(|X_2 y_2|)) > n]$$

$$\leq P[K_2 \max(1,|y_1|^{\gamma_1},|y_2|^{\gamma_2})\max(V_1(|X_1|),V_2(|X_2|)) > n]$$

$$+ P[K_2 \max(1,|y_1|^{\gamma_1},|y_2|^{\gamma_2})\max(V_1(1),V_2(1)) > n]$$

$$\leq K_1(K_2 \max(1,|y_1|^{\gamma_1},|y_2|^{\gamma_2}))^{1+\varepsilon}P^*(n)$$

$$+ K_2 \frac{\max(1,|y_1|^{\gamma_1},|y_2|^{\gamma_2})\max(V_1(1),V_2(1))}{n}$$

for n large enough. Remembering that $\gamma_i = \beta_i/(1+\varepsilon)$ and that $nP^*(n)$ converges,

$$nP[\max(c_1V_1(X_1y_1),c_2V_2(X_2,y_2)) > n] \leq K_3 \max(1,|y_1|^{\beta_1},|y_2|^{\beta_2})$$

for some K_3 and all n large enough.

By dominated convergence and (4.12), therefore,

$$\lim_{n\to\infty} nP[\max(c_1V_1(X_1Y_1),c_2V_2(X_2Y_2)) > n]$$

$$= \int_{\mathcal{R}^2} \lim_{n\to\infty} nP[\max(c_1V_1(X_1y_1),c_2V_2(X_2y_2)) > n]G(dy,dy_2)$$

$$= E\left[\Pi\left(c_1 \frac{[Y_1]^{\alpha_1}}{E|Y_1|^{\alpha_1}}, c_2 \frac{[Y_2]^{\alpha_2}}{E|Y_2|^{\alpha_2}}\right)\right]$$

And this proves $H\varepsilon\mathcal{D}(\alpha_1,\alpha_2)$.

ii) $\alpha_1<\alpha_2=2$. This case is the easiest to prove since marginal convergence is sufficient. Theorem 3.1 yields that H_1, the distribution of X_1Y_1, is in $\mathcal{D}(\alpha_1)$ and Theorem 3.3, part v) tells us that $H_2\varepsilon\mathcal{D}(2)$. Therefore $H\varepsilon\mathcal{D}(\alpha_1,2)$.

iii) $\alpha_1=\alpha_2=2$. We assume $EX_i = EY_i = 0$, without loss of generality. For each i, since $EY_i^2<\infty$, then by Theorem 3.3, X_iY_i has distribution in the domain of attraction of the normal distribution. In fact, we can say more. Let $\mu_i(t) = E[X_i^2 1_{|X_i|\leq t}]$. Since μ_i is slowly varying and is monotone, then we choose K such that for any $t\geq 1$, $u>0$, ε fixed in (0,1),

$$\mu_i(tu) \leq K \max(1,u^\varepsilon)\mu_i(t). \tag{4.14}$$

Therefore, letting G_i be the marginal distribution of Y_i,

$$\begin{aligned} E[(X_iY_i)^2 1_{|X_iY_i|\leq t}] &= \int_{-\infty}^{\infty} y^2\mu_i(t/|y|)G_i(dy) \\ &\leq K \int_{-\infty}^{\infty} y^2 \max(1,|y|^{-\varepsilon})G_i(dy)\mu_i(t) \\ &\leq K\ E[\max(1,Y_i^2)]\mu_i(t) \end{aligned}$$

It follows by dominated convergence that

$$\begin{aligned} \lim_{t\to\infty} \frac{E\left[(X_iY_i)^2 1_{|X_iY_i|\leq t}\right]}{E\left[X_i^2 1_{|X_i|\leq t}\right]} &= \int_{-\infty}^{\infty} y^2 \lim_{t\to\infty} \frac{\mu_i(t/|y|)}{\mu_i(t)} G_i(dy) \\ &= EY_i^2. \end{aligned} \tag{4.15}$$

Let $a_i(t)$ satisfy $t\mu_i(a_i(t))\sim(a_i(t))^2$. Then (4.15) implies

$$\lim_{t\to\infty} \frac{tE\left[(X_iY_i)^2 1_{|X_iY_i|\leq a_i(t)}\right]}{a_i^2(t)} = EY_i^2.$$

This shows that the marginal convergence holds with $a_{in}=a_i(n)$. To show that joint convergence holds, we must consider the truncated covariances. First, define

$$\nu(t,s) = E\left[X_1X_2\, 1_{|X_1|\le t}\, 1_{|X_2|\le s}\right].$$

By assumption,

$$\lim_{t\to\infty} \frac{t\nu\left(\frac{a_1(t)}{u_1}, \frac{a_2(t)}{u_2}\right)}{a_1(t)a_2(t)} = \rho, \qquad \text{for any } u_1>0, u_2>0.$$

Since $(\nu(t_1,t_2))^2 \le \mu_1(t_1)\mu_2(t_2)$, then by (4.14)

$$\left(\nu\left(\frac{t_1}{|y_1|}, \frac{t_2}{|y_2|}\right)\right)^2 \le K^2 \max(1,|y_1|^{-\varepsilon})\max(1,|y_2|^{-\varepsilon})\mu_1(t_1)\mu_2(t_2)$$

and therefore, since $\varepsilon<1$,

$$\left|y_1y_2\nu\left(\frac{a_1(t)}{y_1}, \frac{a_2(t)}{y_2}\right)\right| \le K \max(1,|y_1|)\max(1,|y_2|)\ \frac{a_1(t)a_2(t)}{t}$$

By dominated convergence

$$\lim_{t\to\infty} \frac{tE\left[(X_1Y_1)(X_2Y_2)1_{|X_1Y_1|\le a_1(t)}1_{|X_2Y_2|\le a_2(t)}\right]}{a_1(t)\ a_2(t)}$$

$$= \lim_{t\to\infty}\int_{R^2} y_1y_2\ \frac{t\nu\left(\frac{a_1(t)}{|y_1|}, \frac{a_2(t)}{|y_2|}\right)}{a_1(t)\ a_2(t)}\quad G(dy_1,dy_2)$$

$$= \int_{\mathcal{R}^2} y_1 y_2 \lim_{t\to\infty} \frac{t\nu\left(\dfrac{a_1(t)}{y_1}, \dfrac{a_2(t)}{y_2}\right)}{a_1(t)\ a_2(t)} \ G(dy_1 dy_2)$$

$$= \rho\, EY_1 Y_2 \ . \qquad \#$$

This proves $H \varepsilon \mathcal{D}(2,2)$.

PART IV: REGRESSION WITH INFINITE VARIANCE DATA

Summary. M-estimation is considered for the regression model $Y_j = \beta_0 X_j + W_j$ where X_j's and W_j's are in stable domains of attraction. Necessary and sufficient conditions are given for the consistency of the least squares estimator as well as sufficient conditions for the consistency of other M-estimators. The asymptotic distribution is derived for the least squares estimator. It is found that, depending on the distribution of X_j, different normalization is used and the limit is to a ratio of two jointly distributed stable random variables or to a normal random variable.

1. Introduction

We consider M-estimation for the standard linear regression model (sans intercept)

$$Y_j = \beta_0 X_j + W_j \tag{1.1}$$

where $X_1,\ldots,X_n$ are independent random variables with distribution F and independent of $W_1,\ldots,W_n$ which are independent random variables with distribution G. The observable data is $\{Y_j,X_j\}_{j=1}^{n}$. We assume furthermore that F (in particular) is in a stable domain of attraction.

To be in a domain of attraction, of course, means that normalized partial sums of the random variables converge in distribution. For example let $\mu(t) = \int_{-t}^{t} u^2F(du)$ and choose a_n such that $\frac{1}{a_n^2}\mu(a_n) \sim \frac{1}{n}$ and let $b_n = \int_{-a_n}^{a_n} uF(du)$. We say $F\varepsilon\mathcal{D}(\alpha)$, the domain of attraction of a stable law with index $\alpha\varepsilon(0,2]$, if $\frac{1}{a_n}\sum_{j=1}^{n}(X_j-b_n) \Rightarrow \text{stable}(\alpha)$. For this to be true, it is necessary and sufficient that

$$\lim_{t\to\infty}\frac{\mu(xt)}{\mu(t)} = x^{2-\alpha} \quad \text{for all } x>0 \tag{1.2}$$

and

$$\lim_{t\to\infty}\frac{F(-t)}{1-F(t)+F(-t)} \quad \text{exists when } \alpha<2.$$

The first condition in (1.2) is known as regular variation. When $\alpha<2$, F itself has regular varying tails, namely,

$$\lim_{t\to\infty} \frac{1 - F(xt) + F(-xt)}{1 - F(t) + F(-t)} = x^{-\alpha} \text{ for all } x>0. \tag{1.3}$$

When $\alpha<2$, F does not have finite variance and it may not have finite variance even when $\alpha=2$. In general, if $X\sim F$, then $E|X|^{\beta}<\infty$ for $0<\beta<\alpha$. Our primary concern will be with distributions which have regularly varying tails and infinite variances. Of particular interest is the special case where F and G are tail equivalent (i.e.,

$\lim_{t\to\infty} \frac{1 - F(t) + F(-t)}{1 - G(t) + G(-t)}$ exists and is non zero).

The literature contains much on M-estimation for regression where $X_1,\ldots,X_n$ are fixed, not random (e.g. Huber (1981)). These papers give the motivation for our approach here. Martin and Jong (1977) have considered M-estimation and generalized M-estimation for time series with finite variances. Blattberg and Sargent (1971) and Smith (1973) considered least squares regression with fixed X_j's but stable errors. Kanter and Steiger (1974) investigated least squares and screened ratio estimators for both regression and autogregression under conditions similar to our (1.3). Although we do not consider autoregression here, we view this paper as a stepping stone to that end.

An M-estimator for the true parameter β_0 is $\hat{\beta}_n$ satisfying, for some loss function ρ,

$$L_n(\beta) = \sum_{j=1}^{n} \rho(Y_j - \beta X_j) \text{ is minimum}. \tag{1.4}$$

If ρ is differentiable, we call $\psi = \rho'$ the influence function and $\hat{\beta}_n$ solves the equation

$$K_n(\beta) = \sum_{j=1}^{n} X_j \psi(Y_j - \beta X_j) = 0 \tag{1.5}$$

The idea behind M-estimation, of course, is to mimic procedures in maximum likelihood and least squares estimation, both of which are special cases. Examples of possible loss functions are the following (g is the density of G, the distribution of W_j).

1) least squares $\rho(z) = z^2$ $\psi(z) = 2z$

2) least absolute deviation $\rho(z) = |z|$ $\psi(z) = \mathrm{sgn}(z)$

3) maximum likelihood $\rho(z) = -\ln(g(Z))$ $\psi(z) = -\dfrac{g'(z)}{g(z)}$

4) Cauchy $\rho(z) = \ln(1+z^2)$ $\psi(z) = \dfrac{2z}{1+z^2}$

5) bounded loss $\rho(z) = 1-e^{-\frac{1}{2}z^2}$ $\psi(z) = ze^{-\frac{1}{2}z^2}$

Many others have been suggested by authors, most notably Huber (1981) and Hampel (1971). These authors recommend redescending influence functions, by which they usually mean $\psi(z)=0$ for $|z| \geq c$. This effectively trims the outliers. With contaminated distributions this concept makes sense. However, with regularly varying tails we feel outliers can contribute additional information. If one considers the maximum likelihood estimation when G has regularly varying tails and g' is ultimately monotone, one finds that the influence function ψ satisfies

$\psi(z) = \frac{-g'(z)}{g(z)} \sim \frac{\alpha+1}{z}$ (c.f. Feller, p. 446). This suggests that influence functions with a slow rate of descent, such as $1/z$, are perhaps more appropriate.

In Section 2, we consider the consistency of M-estimators. We first present two results due to Huber (1981), which provide strong consistency when certain expectations exist. We also present a theorem giving necessary and sufficient conditions for the least squares estimator to be weakly consistent.

Section 3 discusses the asymptotic distribution of the least squares estimator. If $F\epsilon\mathcal{D}(\alpha)$, then we will find $c_n(\hat{\beta}_n-\beta_0)$ converges in distribution for normalization c_n, which is regularly varying (as $n\to\infty$) with index $1/\alpha$. In such cases, if $\alpha=2$ the limit distribution is normal, but if $\alpha<2$, it is the ratio of two stable random variables, one with index α and the other with index $\alpha/2$.

The results in Sections 2 and 3 require the joint convergence of

$\left(\frac{1}{a_n}\sum_{j=1}^{k} X_j W_j,\ \frac{1}{a_n^2}\sum_{j=1}^{n} X_j^2\right)$, for which we use Part III.

2. Consistency of M-Estimators

Before we look at consistency for the regression problem, we review consistency for M-estimation in general, following Huber (1981), pp. 127-132. Let $Z_1, Z_2, \ldots$ be a sequence of independent and identically distributed random elements in Z and let $\bar{\rho}(z,\theta)$ be real valued on $Z \times \Theta$. We assume Θ is separable and locally compact with norm $\|\cdot\|$ and has one-point compactification Θ^*.

<u>Lemma 2.1</u> Define $\hat{\theta}_n$ to be any point of minimum in Θ^* of the random function $L_n(\theta) = \sum_{j=1}^{n} \bar{\rho}(Z_j,\theta)$. If $\bar{\rho}$ is almost surely continuous in θ, converges to its maximum as $\|\theta\| \to \infty$, and if θ_0 is such that $\infty > E[\bar{\rho}(Z_j,\theta) - \bar{\rho}(Z_j,\theta_0)] > 0$ for all $\theta \neq \theta_0$, then $\hat{\theta}_n \to \theta_0$ almost surely.

<u>Proof</u>: We note that $\hat{\theta}_n$ exists almost surely in Θ^* because of compactness and continuity. Furthermore, since $\bar{\rho}$ converges to its maximum as $\|\theta\| \to \infty$, then $\hat{\theta}_n$ is almost surely finite. It may not be unique, however. Choose any sequence of solutions.

Define $R_j(\Lambda) = \inf_{\theta' \varepsilon \Lambda} \bar{\rho}(Z_j,\theta') - \bar{\rho}(Z_j,\theta_0)$. For any $\theta \varepsilon \Theta^*$, if $\{\Lambda_i\}$ is a monotone sequence of open sets decreasing to $\{\theta\}, \theta \neq \theta_0$, then by dominated convergence $ER_j(\Lambda_i) \uparrow ER_j(\{\theta\}) > 0$. Hence there exists an open set Λ_θ for each $\theta \varepsilon \Theta^*$ such that $ER_j(\Lambda_\theta) > 0$.

Choose a covering $\Lambda_1,\ldots,\Lambda_k$ for the region $\|\theta-\theta_0\|\geq\varepsilon$, which is compact in Θ^*. For each $i=1,\ldots,k$

$$
\begin{aligned}
&P[\hat{\theta}_n\varepsilon\Lambda_i \text{ for infinitely many values of } n]\\
&\qquad \leq P[L_n(\theta) \text{ is minimum for some } \theta\varepsilon\Lambda_i \text{ i.0.}]\\
&\qquad \leq P[\inf_{\theta\varepsilon\Lambda_i} L_n(\theta) \leq L_n(\theta_0) \text{ i.o.}]\\
&\qquad \leq P\left[\sum_{j=1}^{n} R_j(\Lambda_i) \leq 0 \text{ i.o.}\right]\\
&\qquad = 0 .
\end{aligned}
$$

The last equality holds because $\frac{1}{n}\sum_{j=1}^{n} R_j(\Lambda_i) \to ER_j(\Lambda_i) > 0$ almost surely.

Therefore, we have

$$
\begin{aligned}
P[\|\hat{\theta}_n-\theta_0\|\geq\varepsilon \text{ i.o.}] &\leq \sum_{i=1}^{k} P[\hat{\theta}_n\varepsilon\Lambda_i \text{ i.o.}]\\
&= 0.
\end{aligned}
$$

#

Alternatively, the M-estimator can be defined by the use of implicit equations and the next lemma can be used to verify consistency. We assume that the equation $\bar{\psi}(Z_1,\theta) = 0$ has a solution almost surely.

<u>Lemma 2.2</u> Let $\hat{\theta}_n$ be a root of the function $K_n(\theta) = \sum_{j=1}^{n}\bar{\psi}(Z_j,\theta)$ and suppose that $E|\bar{\rho}(Z_1,\theta)|<\infty$ for all $\theta\varepsilon\Theta$ and $E\bar{\psi}(Z_1,\theta)$ has a unique root θ_0 and is monotone in a neighborhood of θ_0. Let

$$S_j(\Lambda,\theta) = \sup_{\theta'\varepsilon\Lambda}|\bar{\psi}(Z_j,\theta') - \bar{\psi}(Z_j,\theta)| .$$

If, as the sets $\Lambda_i\downarrow\{\theta\}$, $ES_j(\Lambda_i,\theta)\downarrow 0$ for all $\theta\varepsilon\Theta^*$, then $\hat{\theta}_n\to\theta_0$ almost surely.

For the proof of this, see Huber (1964). We remark that the condition $ES_j(\Lambda_i,\theta)\downarrow 0$ is easily satisfied when $\bar{\psi}$ is continuous in θ and either $|\bar{\psi}(Z_j;\theta)| \leq K(Z_j)$ where $EK(Z_j)<\infty$ or θ is real and $\bar{\psi}$ is monotone in θ.

In our regression problem, the data is given by $\{Z_j\}_{j=1}^{n} = \{(Y_j,X_j)\}_{j=1}^{n}$ and $\bar{\rho}(Y_j, X_j,\beta) = \rho(Y_j-\beta X_j)$ where the latter function has a real argument. We also have $\bar{\psi}(Y_j,X_j,\beta) = X_j\psi(Y_j-\beta X_j)$, when $\psi=\rho'$ exists.

<u>Corollary 2.3</u> a) Suppose $\hat{\beta}_n$ satisfies (1.4) where $\rho(z)$ is continuous and converges to its maximum as $|z|\to\infty$. Assume $E[\rho(W_j-\theta X_j)-\rho(W_j)]<\infty$ for all $\theta\varepsilon R$. If $E[\rho(W_j-\theta X_j)-\rho(W_j)]$ has a unique minimum at $\theta=0$, then $\hat{\beta}_n\to\beta_0$ almost surely.

b) Suppose $\hat{\beta}_n$ satisfies (1.5) where ψ is continuous and either monotone or bounded. Assume $E|X_j\,\psi(W_j-\theta X_j)|<\infty$ for all $\theta\varepsilon R$. If $\theta=0$ is the unique root of $E[X_j\,\psi(W_j-\theta X_j)]$, then $\hat{\beta}_n\to\beta_0$ almost surely.

<u>Proof</u>: The proof is a direct application of Lemma 2.1 or Lemma 2.2 when we recognize that $Y_j-\beta X_j = W_1-(\beta-\beta_0)X_j$ and substitute θ for $\beta-\beta_0$. #

The crucial condition in Corollary 2.3 is the unique minimum (root). When $\rho=-\ln g$ so that $\tilde{\beta}_n$ is the maximum lokelihood estimator, Jensen's inequality assures us this condition is satisfied. We can also satisfy the condition by strengthening conditions on the distribution of W and on the loss function, but with very little assumption on the distribution of X.

<u>Lemma 2.4</u> Suppose W has distribution G which is symmetric about 0 and strictly increasing at 0 and suppose X has distribution F, not degenerate at 0.

a) Let ρ be a continuous function, symmetric about a unique minimum at 0 and nondecreasing on $(0,\infty)$. Assume $E|\rho(W_j-\theta X_j)-\rho(W_j)|<\infty$ for all θ. If either ρ is convex or G is unimodal, then $E|\rho(W_j-\theta X_j)-\rho(W_j)|$ is continuous, symmetric about a unique minimum at $\theta=0$ and nondecreasing on $(0,\infty)$.

b) Let ψ be a continuous function, antisymmetric about a unique root at 0 and nonnegative on $(0,\infty)$. Assume $E|X_j\psi(W_j-\theta X_j)|<\infty$ for all θ. If either ψ is monotone or G is unimodal, then $E[X_j\psi(W_j-\theta X_j)]$ is continuous, antisymmetric about a unique root at $\theta=0$ and nonnegative on $(0,\infty)$.

<u>Proof</u>: We prove only a) since b) has a very similar proof. Clearly if G is degenerate at 0, the condition holds since $P[\rho(\theta X_j)>\rho(0)]>0$ when $\theta\neq 0$. We assume therefore that G is not degenerate at 0. We can also assume without loss of generality that $\rho(0)=0$. Let

$$\begin{aligned}\ell(y) &= E[\rho(W_j-y) - \rho(W_j)] \\ &= \int_{-\infty}^{\infty} [\rho(w-y) - \rho(w)]G(dw) \qquad (3.1)\end{aligned}$$

Clearly, ℓ is continuous since ρ is continuous. By the symmetry in ρ and G,

$$\ell(y) = \int_0^{\infty} [\rho(w-y) - \rho(w) + \rho(-w-y) - \rho(-w)]G(dw)$$

$$= \int_0^\infty [\rho(w+y) + \rho(w-y) - 2\rho(w)]G(dw). \tag{3.2}$$

From this it is apparent that $\ell(y) = \ell(-y)$.

Suppose ρ is convex. Then ρ must be strictly increasing on $[0,\infty)$. Otherwise, if for some $0<y<z$, $\rho(y) = \rho(z)$ then setting $\lambda=y/z$ we have

$$\begin{aligned}\rho(y) &> \lambda\rho(z) + (1-\lambda)\rho(0)\\ &\geq \rho(y),\end{aligned}$$

a contradiction. (Of course, we also cannot have $\rho(y) = \rho(0) = 0$, since 0 is ρ's unique minimum.) Therefore for any $y>2w\geq 0$

$$\rho(y+w) + \rho(y-w) - 2\rho(w) > 0 .$$

On the other hand, since ρ is convex,

$$\rho(y+w) + \rho(y-w) - 2\rho(w) \geq 0$$

for any $y\geq 0$, $w\geq 0$. Using (3.2) and the fact that G is strictly increasing at zero,

$$\begin{aligned}\ell(y) &= \int_0^\infty [\rho(y+w) + \rho(y-w) - 2\rho(w)]G(dw)\\ &\geq \int_0^{y/2} [\rho(y+w) + \rho(y-w) - 2\rho(w)]G(dw)\\ &> 0 .\end{aligned}$$

Furthermore, convexity implies for $z>y\geq 0$ $\left(\lambda = \frac{z+y}{2z}\right)$

$$\rho(w+z) + \rho(w-z) - \rho(w+y) - \rho(w-y)$$

$$= \lambda\rho(w+z) + (1-\lambda)\rho(w-z) - \rho(w+\lambda z-(1-\lambda)z)$$

$$+ (1-\lambda)\rho(w+z) + \lambda\rho(w-z) - \rho(w+(1-\lambda)z-\lambda z)$$

$$\geq 0 \ . \tag{3.3}$$

And since ρ is increasing on $(0,\infty)$, (3.3) is a strict inequality when $0\leq w\leq y$. Therefore, from (3.2), if $z>y>0$

$$\ell(z)-\ell(y) = \int_0^\infty [\rho(w+z)+\rho(w-z)-\rho(w+y)-\rho(w-y)]G(dw)$$

$$\geq \int_0^y [\rho(w+z)+\rho(w-z)-\rho(w+y)-\rho(w-y)]G(dw)$$

$$> 0 \ .$$

Thus $\ell(y)$ is increasing on $(0,\infty)$ and symmetric about a unique minimum at 0.

Suppose instead that G is unimodal. Then (c.f. Feller (1971), p. 158) G has a density g and a possible mass p at 0 and the density can be characterized by

$$g(w) = \int_{u\geq|w|} \frac{1}{2u} H(du)$$

for some probability measure H. Since G is increasing at 0, then either $p>0$ or H has mass in any neighborhood of 0.

Since $|\rho(w-y)-\rho(w)|$ is integrable with respect to g and g is itself an integral with positive integrand, we use Fubini's theorem,

$$\ell(y) = \int_{-\infty}^{\infty} [\rho(w-y)-\rho(w)]G(dw)$$

$$= \int_{-\infty}^{\infty} [\rho(w-y)-\rho(w)] \int_{u\geq|w|} \frac{1}{2u} H(du)dw + p\rho(y)$$

$$= \int_{0}^{\infty} \int_{-u}^{u} [\rho(w-y)-\rho(w)]dw\ H(du) + p\rho(y)$$

$$= \int_{0}^{\infty} \left[\int_{-u-y}^{u-y} \rho(w)dw - \int_{-u}^{u} \rho(w)dw \right] H(du) + p\rho(y)$$

$$= \int_{0}^{\infty} \left[\int_{-u-y}^{-u} \rho(w)dw - \int_{u-y}^{u} \rho(w)dw \right] H(du) + p\rho(y)$$

$$= \int_{0}^{\infty} \int_{-y}^{0} [\rho(w-u)-\rho(w+u)]dw\ H(du) + p\rho(y).$$

By the symmetry of ρ,

$$\ell(y) = \int_{0}^{\infty} \int_{0}^{y} [\rho(w+u)-\rho(w-u)]dw\ H(du) + p\rho(y). \tag{3.4}$$

Since ρ is nondecreasing on $(0,\infty)$ and symmetric about 0, then for any $w,u\geq 0$

$$\rho(w+u) - \rho(w-u) \geq 0 \tag{3.5}$$

and therefore, if $0\leq y<z$, we obtain from (3.4)

$$\ell(z)-\ell(y) = \int_{0}^{\infty} \int_{y}^{z} [\rho(w+u)-\rho(w-u)]dw\ H(du) + p[\rho(z)-\rho(y)]$$

$$\geq 0.$$

Furthermore, ρ is strictly increasing near zero, so that (3.5) is a strict inequality when $0<u\le c$, $0\le w\le c$, for some $c>0$. Since either $p>0$ or H is increasing at 0, then

$$\begin{aligned}\ell(y) &= \int_0^\infty \int_0^y [\rho(w+u)-\rho(w-u)]dw\, H(du) + p\rho(y) \\ &\ge \int_0^c \int_0^y [\rho(w+u)-\rho(w-u)]dw\, H(du) + p\rho(y) \\ &> 0\ .\end{aligned}$$

Again, therefore, ℓ is symmetric about a unique minimum at 0 and nondecreasing on $(0,\infty)$.

Finally, we write $E[\rho(W_j - \theta X_j)-\rho(W_j)] = E[\ell(\theta X_j)]$ and since $P[X=0]<1$, it is apparent that this is also symmetric about a unique minimum at 0 and nondecreasing on $(0,\infty)$. #

As a corollary, the conditions of Lemma 2.4 imply that the corresponding M-estimator is strongly consistent. The only condition placed on F is that it is not degenerate at 0 and that $E|\rho(W_j-\theta X_j) - \rho(W_j)|<\infty$ $(E|X_j\psi(W_j-\theta X_j)|<\infty)$. This means that, under the conditions in Lemma 2.4, Corollary 2.3 can easily be extended to include location estimation and multiple regression.

In addition to the unique minimum (root) condition in Corollary 2.3, the other restrictive condition is the expectation. If F has infinite variance, for example, we cannot show that the least squares estimator is strongly consistent. We shall see below, however, that it can still be weakly consistent. Before that, we summarize our results so far for distributions with regularly varying tails.

Theorem 2.5 Suppose $X_j \sim F$ and $W_j \sim G$ such that $E|X_j|^{\alpha_1} < \infty$, $E|W_j|^{\alpha_2} < \infty$ for some $\alpha_1 > 0$, $\alpha_2 > 0$. Assume also G is symmetric about 0 and increasing at 0. Choose $\hat{\beta}_n$ to satisfy (1.4), where $\rho(w)$ is continuous, symmetric about a unique minimum at 0 and nondecreasing on $(0,\infty)$. Assume also that $\rho(w)$ is regularly varying (as $w\to\infty$) with exponent $\gamma\varepsilon[0, \min(\alpha_1,\alpha_2))$. Then $\hat{\beta}_n \to \beta_0$ almost surely if either ρ is convex (which requires $\gamma \geq 1$) or G is unimodal. If in fact $\rho(w) = kw^{\gamma}$ for $w \geq M$, then it is sufficient that $\gamma < \alpha_1$, $\frac{\gamma(j-1)}{j} < \alpha_2$ where j= integer in $[\gamma,\gamma+1)$.

Proof: In view of Corollary 2.3 and Lemma 2.4, the only condition requiring proof is that $E|\rho(W_j-\theta X_j)-\rho(W_j)| < \infty$ for all θ, since if that is true, Lemma 2.4 gives $E[\rho(W_j-\theta X_j)-\rho(W_j)]$ has a unique minimum at $\theta=0$ and Corollary 2.3 yields the consistency result.

Since $\gamma < \min(\alpha_1,\alpha_2)$, choose $\delta = \min(\alpha_1,\alpha_2)$. Then $E|X_j|^{\delta} < \infty$ and $E|W_j|^{\delta} < \infty$.

Since ρ is regularly varying, it has Karamata representation (c.f. Feller (1971), p. 282)

$$\rho(w) = c(w)\exp\left[\int_1^w \frac{\gamma(u)}{u}\,du\right]$$

where $c(w)\to$constant and $\gamma(w)\to\gamma$, as $w\to\infty$. Choose w_0 such that $\sup_{w\geq w_0} \gamma(w) \leq \delta$ and let $K_1 = \sup_{w\geq w_0} \frac{c(w)}{c(w_0)} \frac{\rho(w_0)}{w_0^{\delta}}$. Therefore, if $w \geq w_0$ then

$$\rho(w) = c(w)\exp\left[\int_1^{w_0} \frac{\gamma(u)}{u}\,du\right] \exp\left[\int_{w_0}^{w} \frac{\delta}{u}\,du\right]$$

$$\leq K_1 w^{\delta}$$

Using ρ's symmetry and monotonicity on $(0,\infty)$,

$$\rho(w) \leq K_1 \max(1, |w|^{\delta})$$

for all $w\varepsilon R$. Therefore, for any $w\varepsilon R$, $x\varepsilon R$

$$\rho(w+x) \leq K_1 \max(1, |w+x|^{\delta})$$

$$\leq K_2 \max(1, |w|^{\delta}+|x|^{\delta})$$

for some constant K_2. Hence

$$E[\rho(W_j - \theta X_j)] \leq K_2 E[\max(1, |W_j|^{\delta} + |X_j|^{\delta})] < \infty.$$

This guarantees that $E|\rho(W_j+\theta X_j)-\rho(W_j)|<\infty$.

If we can assume further that $\rho(w)=k|w|^{\gamma}$, $|w|\geq M$ for some k,M, then we need only show that $E[|W_j+\theta X_j|^{\gamma}-|W_j|^{\gamma}]<\infty$. Let j=integer in $[\gamma,\gamma+1)$. By Minkowski's inequality, $w\geq 1, x\geq 1$

$$(w+x)^{\gamma}-w^{\gamma} = ((w+x)^{\gamma/j})^{j} - (w^{\gamma/j})^{j}$$

$$\leq (w^{\gamma/j}+x^{\gamma/j})^{j} - (w^{\gamma/j})^{j}$$

$$= \sum_{i=0}^{j-1} \binom{j}{i} w^{\gamma i/j} x^{\gamma(j-i)/j}$$

$$\leq K w^{\gamma(j-1/j)} x^{\gamma}$$

for some constant K. In general, for any w,x

$$|w+x|^{\gamma}-|w|^{\gamma} \le K\max(1,|w|^{\gamma(j-1/j)})\max(1,|x|^{\gamma}).$$

Therefore $E[|W_j+\theta X_j|^{\gamma}-|W_j|^{\gamma}]<\infty$ if $\gamma<\alpha_1$ and $\gamma<\frac{j\alpha_2}{j-1}$, $\gamma\le j$. #

Theorem 2.6 Assume F and G are as in Theorem 2.5 ($\alpha_1>1$) and choose $\hat{\beta}_n$ to satisfy (1.5) for the model (1.1), where $\psi(w)$ is continuous, antisymmetric about 0, monotone and regularly varying as $W\to\infty$ with exponent $\gamma\ge 0$, $\gamma<\min(\alpha_1-1,\alpha_2)$, then $\hat{\beta}_n\to\beta_0$ almost surely.

Proof: Relying on Lemma 2.4 we see that if $E[X_j\psi(W_j-\theta X_j)]$ exists, then it has a unique root at $\theta=0$ and so by Corollary 2.3, $\hat{\beta}_n\to\beta_0$ almost surely. However, as in the previous theorem, due to the regular variation of ψ, we can write, for any $\delta>\gamma$, some $K_1>0$

$$\begin{aligned}|\psi(w)| &= \psi(|w|)\\ &\le K_1\max(1,|w|^{\delta}).\end{aligned}$$

Choosing $\delta=\min(\alpha_1-1,\alpha_2)$, we have therefore for some K_2,

$$\begin{aligned}E|X_j\psi(W_j-\theta X_j)| &\le K_1E[|X_j|\max(1,|W_j-\theta X_j|^{\delta})]\\ &\le K_2E[|X_j|\max(1,|W_j|^{\delta}+|\theta X_j|^{\delta})]\\ &< \infty. \qquad \#\end{aligned}$$

We consider now the special example of the least squares estimator, where $\rho(w)=w^2$. Since $\rho(W_j-\theta X_j)-\rho(W_j) = \theta^2X_j^2 - 2\theta X_jW_j$, then by Corollary 2.3, the estimator will be strongly consistent when $EX_j^2<\infty$, $EW_jX_j=0$. The next theorem shows that it is weakly consistent for a great many more cases. We make no assumptions of symmetry. The idea for the theorem comes from Kanter and Steiger (1974) who prove a special case.

Theorem 2.7 Suppose F and G are in domains of attraction ($F\varepsilon\mathcal{D}(\alpha_1)$, $G\varepsilon\mathcal{D}(\alpha_2)$, $\alpha_1,\alpha_2 \leq 2$) and define $\mu(t) = \int_{-t}^{t} x^2 F(dx)$. For the regression model (1.1), define the least squares estimator $\hat{\beta}_n = \dfrac{\sum_{j=1}^{n} X_j Y_j}{\sum_{j=1}^{n} X_j^2}$. We consider three cases.

i) $\alpha_1 < 2\alpha_2$. $\hat{\beta}_n \to \beta_0$ in probability, except if $EW_jX_j \neq 0$ and $EX_j^2 < \infty$.

ii) $\alpha_1 > 2\alpha_2$. $\hat{\beta}_n$ is not consistent.

iii) $\alpha_1 = 2\alpha_2$. $\hat{\beta}_n \to \beta_0$ in probability if and only if

$$\lim_{t\to\infty} \frac{t^2 P[|W_j| > t^2]}{\mu(t)} = 0 \text{ and, when } \alpha_1 = 2,$$

$$\lim_{n\to\infty} \frac{E\left[W_j \, 1_{|W_j| \leq c_n}\right]}{\mu(a_n)} = 0, \text{ where } c_n \text{ and } a_n \text{ are given in the proof.}$$

Proof: Since $F\varepsilon\mathcal{D}(\alpha_1)$, then by Feller (1971), p. 577, μ is regularly varying with exponent $2-\alpha_1$. Define $A(t) = \frac{1}{t^2}\mu(t)$ and $a_n = A^{\leftarrow}(1/n)$ $= \inf(x{:}A(x) \leq 1/n)$. We note that $\lim_{n\to\infty} \frac{n}{a_n}\mu(a_n) = 1$. When $\alpha_1 < 2$, then the distribution of X^2 has $\alpha_1/2$-varying tails and in fact is in $\mathcal{D}(\alpha_1/2)$. Thus $\frac{1}{a_n^2}\sum_{j=1}^{n} X_j^2 \Rightarrow$ positive stable $(\alpha_1/2)$. When $\alpha_1 = 2$, then μ is slowly varying and so by Feller (1971), p. 236, $\frac{1}{a_n^2}\sum_{j=1}^n X_j^2 \to 1$ in probability.

Recognizing that $\hat{\beta}_n - \beta_0 = \frac{\sum_{j=1}^n X_j W_j}{\sum_{j=1}^n X_j^2}$, we conclude that $\hat{\beta}_n \to \beta_0$ in probability if and only if $\frac{1}{a_n^2}\sum_{j=1}^n X_j W_j \to 0$ in probability.

Let H be the distribution of $|X_j W_j|$ and let $\alpha^*=\min(\alpha_1,\alpha_2)$. By Corollary 3.2 and Theorem 3.3 of Part III, $H\varepsilon\mathcal{D}(\alpha^*)$. This means if we define $C(t) = \frac{1}{t^2}\int_0^t u^2 H(du)$ and c_n and b_n to satisfy

$$c_n = C^{\leftarrow}(1/n)$$

and

$$b_n = \begin{cases} 0 \text{ if } \alpha^*<1 \\ \int_0^{C_n} u\, H(du) \text{ if } \alpha^*\geq 1, \end{cases}$$

then $\frac{1}{c_n}\sum_{j=1}^n (|X_j W_j| - b_n) \Rightarrow$ stable (α^*). We note that C(t) is regularly varying with exponent $-\alpha^*$ and thus, by deHaan (1970) p. 22, c_n is regularly varying with exponent $1/\alpha^*$. Similarly, a_n is regularly varying with exponent $1/\alpha_1$. The sequence b_n is either slowly varying ($a^*\geq 1$) or identically zero ($\alpha^*<1$).

i) Suppose $\alpha_1<2\alpha_2$. Assume first $EX_j^2=\infty$. Since $\alpha^*=\min(\alpha_1,\alpha_2)>\alpha_1/2$, then $\frac{c_n}{a_n^2}\to 0$. If $\alpha_1<2$, then either $nb_n=0$ or nb_n is regularly varying with exponent 1. Since $2/\alpha_1>1$, then $\frac{nb_n}{a_n^2}\to 0$. On the other hand, if $\alpha_1=2$, then $\frac{n}{a_n^2}\sim\frac{1}{\mu(a_n)}\to\frac{1}{EX_j^2}=0$ and $b_n\to E|W_jX_j|$ (since $\alpha_2>1$ in this case).

Again $\frac{nb_n}{a_n^2}\to 0$. We therefore have

$$\frac{1}{a_n^2}\sum_{j=1}^{n}X_jW_j \le \frac{1}{a_n^2}\sum_{j=1}^{n}|X_jW_j|$$

$$= \frac{c_n}{a_n^2}\left(\frac{1}{c_n}\sum_{j=1}^{n}(|X_jW_j|-b_n)\right)+\frac{nb_n}{a_n^2}$$

$$\to 0 \text{ in probability.}$$

Thus $\hat{\beta}_n\to\beta_0$ in probability when $\alpha_1<2\alpha_2$, $EX_j^2=\infty$.

If $\alpha_1<2\alpha_2$, $EX_j^2<\infty$, then $\lim_{n\to\infty}\frac{a_n^2}{n}=\lim_{n\to\infty}\mu(a_n)=EX_j^2$. Also $\alpha_2>1$ so that $\frac{1}{n}\sum_{j=1}^{n}X_jW_j\to EX_jW_j$ almost surely. Thus, $\frac{1}{a_n^2}\sum_{j=1}^{n}X_jW_j\to\frac{EX_jW_j}{EX_j^2}$ almost surely and $\hat{\beta}_n$ is strongly consistent if and only if $EX_jW_j=0$.

ii) $\alpha_1>2\alpha_2$. In this case $\alpha^*=\min(\alpha_1,\alpha_2)=\alpha_2<\alpha_1/2$. Therefore $\frac{c_n}{a_n^2}\to\infty$. It is thus impossible for $\frac{1}{a_n^2}\sum_{j=1}^{n}X_jW_j$ to converge to 0, since $\frac{1}{c_n}\sum_{j=1}^{n}(|X_jW_j|-b_n)$ converges in distribution. Hence $\hat{\beta}_n$ is not consistent.

iii) $\alpha_1=2\alpha_2$. Here, $\alpha^*=\alpha_2\le 1$. In this case the sequence $\frac{c_n}{a_n^2}$ is slowly varying. By Corollary 3.2, Part III, the distribution of X_jW_j is also in $\mathcal{D}(\alpha^*)$ with the same normalization c_n as for $|X_jW_j|$. Define

$$d_n = \begin{cases} 0 \text{ if } \alpha^*<1 \\ E\left[W_j\, 1_{|W_j|\le c_n}\right]EX_j \quad \text{if } \alpha^*=1. \end{cases}$$

Then (Feller (1971), p. 580), $\frac{1}{c_n}\sum_{j=1}^{n}(X_jW_j-d_n) \Rightarrow$ stable (1). It follows, therefore, that $\frac{1}{a_n^2}\sum_{j=1}^{n} X_jW_j \to 0$ if and only if $\frac{c_n}{a_n^2}\to 0$ and $\frac{nd_n}{a_n^2}\to 0$. We show below that the first condition is equivalent to $\frac{t^2P[|W_j|>t^2]}{\mu(t)} \to 0$. When $\alpha^*<1$, $d_n=0$ and the second condition is automatic. Therefore, when $\alpha_1=2\alpha_2<2$, $\hat{\beta}_n\to\beta_0$ in probability if and only if $\lim_{t\to\infty}\frac{t^2P[|W_j|>t^2]}{\mu(t)}=0$. When $\alpha^*=1$, however, the condition $\frac{nd_n}{a_n^2}\to 0$ must be checked for each case.

Noting that $\frac{n\mu(a_n)}{a_n^2}\to 1$, we state: when $\alpha_1=2\alpha_2=2$, $\hat{\beta}_n\to\beta_0$ in probability if and only if $\lim_{t\to\infty}\frac{t^2P[|W_j|>t^2]}{\mu(t)} = \lim_{n\to\infty}\frac{E[W_j\, 1_{|W_j|\le c_n}]}{\mu(a_n)} = 0.$

It remains only to show that $\frac{c_n}{a_n^2}\to 0$ if and only if $\frac{t^2P[|W_j|>t^2]}{\mu(t)}\to 0$. Recall that $c_n = C^{\leftarrow}(1/n)$ and $a_n = A^{\leftarrow}(1/n)$. Since C is regularly varying with exponent $-\alpha^*$ (by definition of C) and A is regularly varying with exponent $-\alpha_1$, then we apply the result in deHaan (1970), p. 22, which says $\frac{c_n}{a_n^2} \to 0$ if and only if $\frac{C(t)}{A(t^{\frac{1}{2}})} \to 0$. (That is, the inverses converge as one would expect, for monotone, regularly varying functions.) Furthermore, by Feller (1971), p. 283,

$$\lim_{t\to\infty} \frac{1-H(t)}{C(t)} = \frac{2-\alpha^*}{\alpha^*} = \frac{2-\alpha_2}{\alpha_2}$$

and by Lemma 2.1, Part III,

$$\lim_{t\to\infty} \frac{1-H(t)}{P[|W_j|>t]} = E|X_j|^{\alpha_2} .$$

Therefore

$$\lim_{t\to\infty} \frac{C(t)}{A(t^{\frac{1}{2}})} = \left(\frac{\alpha_2 E|X_j|^{\alpha_2}}{2-\alpha_2}\right) \lim_{t\to\infty} \frac{P[|W_j|>t]}{A(t^{\frac{1}{2}})}$$

$$= \left(\frac{\alpha_2 E|X_j|^{\alpha_2}}{2-\alpha_2}\right) \lim_{t\to\infty} \frac{t^2 P[|W_j|>t^2]}{\mu(t)}$$

It follows, then, that $\frac{c_n}{a_n^2} \to 0$ if and only if $\lim_{t\to\infty} \frac{t^2 P[|W_j|>t^2]}{\mu(t)} = 0$. #

When $\alpha_1=2\alpha_2=2$, the condition $\frac{E\left[W_j 1_{|W_j|\leq c_n}\right]}{\mu(an)} \to 0$ is satisfied when the numerator vanishes (e.g., G is symmetric about 0 or $EW_j=0$) or when the numerator is bounded and $\mu(a_n)\to\infty$ (e.g., $E|W_j|<\infty$ and $EX_j^2=\infty$). The condition $\dfrac{t^2 P[|W_j|>t^2]}{\mu(t)} \to 0$ holds for case i) and is false for case ii).

It is thus the primary condition for least squares consistency; the centering conditions are necessary only when $\alpha_1 = 2$.

3. Asymptotic Distribution of Least Squares Estimators

In this section, we prove a theorem describing the asymptotic distribution of the least squares estimator. For simplicity, symmetric distributions are assumed.

Theorem 3.1 Suppose $X_j \sim F \in \mathcal{D}(\alpha)$ and $W_j \sim G$ where F and G are symmetric about 0. Let $\hat{\beta}_n$ be the least squares estimator for the model (1.1). Choose a_n such that

$$\lim_{n\to\infty} \frac{n}{a_n^2} E[X_j^2 1_{|X_j| \le a_n}] = 1. \quad \text{Then}$$

i) if $\alpha<2$ and $E|W|^{\gamma}<\infty$ for some $\gamma > \alpha$, $a_n(\hat{\beta}_n - \beta_0) \Rightarrow Z_1/Z_2$ where (Z_1, Z_2) is bivariate stable and Z_1 is symmetric stable (α_1), Z_2 is positive stable $(\alpha/2)$. (Their joint distribution has Levy measure ν given by, $c_1 c_2 \neq 0$,

$$\nu\{\max(c_1[x_1]^{1/\alpha}, c_2[x_2]^{2/\alpha})>1\} = \tfrac{1}{2}E \max |c_1| \frac{|W_j|^{\alpha}}{E|W_j|^{\alpha}}, c_2$$

$$= \tfrac{1}{2}E\left[\max\left(|c_1|\frac{|W_j|^{\alpha}}{E|W_j|^{\alpha}}, c_2\right)\right] + \tfrac{1}{2}\max(0, c_2).)$$

and

ii) if $\alpha=2$ and $EW_j^2<\infty$, $a_n(\hat{\beta}_n - \beta_0) \Rightarrow$ normal $(0, EW_j^2)$.

Proof: i) By assumption, $F\varepsilon\mathcal{D}(\alpha)$, $\alpha<2$ and F is symmetric, so that

$$\lim_{t\to\infty} \frac{P[|X_j|>xt]}{P[|X_j|>t]} = x^{-\alpha}, \quad x>0, \tag{3.1}$$

and

$$\lim_{t\to\infty} \frac{P[X_j>t]}{P[|X_j|>t]} = \tfrac{1}{2} . \tag{3.2}$$

(3.1) implies

$$\lim_{t\to\infty} \frac{P[X_j^2>xt]}{P[X_j^2>t]} = x^{-\alpha/2}, \quad x>0.$$

And this in turn implies that the distribution of X_j^2 is in the domain of attraction of a positive stable ($\alpha/2$). We therefore have (Feller (1971), p. 580)

$$\frac{1}{a_n} \sum_{j=1}^{n} X_j \Rightarrow \text{symmetric stable } (\alpha)$$

$$\frac{1}{a_n^2} \sum_{j=1}^{n} X_j^2 \Rightarrow \text{positive stable } (\alpha/2)$$

(No centerings are needed by symmetry and the fact that $\alpha/2<1$.)

To check that these in fact converge jointly, we use Lemma 4.1 of Part III. Define for $x\varepsilon\mathcal{R}$, $i=1,2$,

$$U_i(x) = \frac{\operatorname{sgn}(x)}{P[|X_j^i|>x]} \tag{3.3}$$

Clearly $U_2(x) = U_1(\operatorname{sgn}(x)|x|^{\frac{1}{2}})$, and therefore for fixed $c_1, c_2 \varepsilon\mathcal{R}$,

$$\max(c_1U_1(X_j), c_2U_2(X_j^2)) = \max(c_1U_1(X_1), c_2U_1(|X_j|)). \tag{3.4}$$

From (3.2) and (3.3) we calculate

$$\lim_{t\to\infty} tP[U_1(X_j)>t] = \lim_{t\to\infty} tP[X_j>U_1^{\leftarrow}(t)]$$

$$= \lim_{s\to\infty} U_1(s)P[X_j>s]$$

$$= \lim_{s\to\infty} \frac{P[X_j>s]}{P[|X_j|>s]}$$

$$= \tfrac{1}{2} .$$

Similarly $\lim_{t\to\infty} tP[-U_1(X_j)>t] = \frac{1}{2}$. And therefore, for any $c\in\mathcal{R}$,

$$\lim_{t\to\infty} tP[c\, U_1(X_j)>t] = \tfrac{1}{2}|c| . \tag{3.5}$$

Using (3.4) and (3.5),

$$\lim_{t\to\infty} tP[\max(c_1U_1(X_j),c_2U_2(X_j^2))>t]$$

$$= \lim_{t\to\infty} tP[\max(c_1U_1(X_j),c_2U_1(|X_j|))>t]$$

$$= \lim_{t\to\infty} t\Big[P[\max(c_1,c_2,0)U_1(X_1)>t]$$

$$+ P[\min(c_1,-c_2,0)U_1(X_1)>t]\Big]$$

$$= \tfrac{1}{2}|\max(c_1,c_2,0)|-\tfrac{1}{2}|\min(c_1,c_2,0)|$$

$$= \tfrac{1}{2}\max(|c_1|,c_2) + \tfrac{1}{2}\max(0,c_2)$$

$$= \Pi(c_1,c_2) \tag{3.6}$$

And this satisfies the condition required in Lemma 4.1, Part III for joint convergence.

Since $E|W_j|^\gamma<\infty$ for some $\gamma>\alpha$, then by Theorem 4.2 of Part III, the distribution of (X_jW_j, X_j^2) is also in a bivariate domain of attraction. From the proof of that theorem, it is apparent that the same normalizations, a_n and a_n^2, are appropriate. Again, no centering is required. Therefore, we have

$$\left(\frac{1}{a_n}\sum_{j=1}^n X_jW_j, \ \frac{1}{a_n^2}\sum_{j=1}^n X_j^2\right) \Rightarrow (Z_1, Z_2)$$

where (Z_1,Z_2) is jointly stable and $Z_1 \sim$ symmetric stable (α) and $Z_2 \sim$ positive stable $(\alpha/2)$.

By the continuous mapping theorem (and since $Z_2>0$ almost surely),

$$a_n(\hat{\beta}_n-\beta_0) = \frac{\frac{1}{a_n}\sum_{j=1}^n X_jW_j}{\frac{1}{a_n^2}\sum_{j=1}^n X_j^2} \Rightarrow Z_1/Z_2.$$

The joint distribution of (Z_1,Z_2) can be expressed in terms of its Levy measure ν, which from (3.6) and the proof of Theorem 4.2, Part III is determined by (letting $[x]^\gamma = \mathrm{sgn}(x)|x|^\gamma$)

$$\nu\{\max(c_1[x_1]^{1/\alpha}, c_2[x_2]^{2/\alpha})>1\} = E\left[\Pi\left(c_1\frac{[W_j]^\alpha}{E|W_j|^\alpha}, c_2\right)\right]$$

$$= \tfrac{1}{2}E\left[\max\left(|c_1|\frac{|W_j|^\alpha}{E|W_j|^\alpha}, c_2\right)\right] + \tfrac{1}{2}\max(0,c_2).$$

ii) When $F\varepsilon\mathcal{D}(2)$, then $E[X_j^2 1_{|X_j|\leq t}]$ is slowly varying as $t\to\infty$, so that Feller (1971), p. 236, yields

$$\frac{1}{a_n^2} \sum_{j=1} X_j^2 \to 1 \text{ in probability.}$$

We also have that the distribution of X_jW_j is in $\mathcal{D}(2)$, by Theorem 3.3, Part III, and again the normalization a_n is appropriate, since $EW_j^2<\infty$. Thus, by Feller (1971), p. 580,

$$\frac{1}{a_n} \sum_{j=1}^{n} X_jW_j \Rightarrow \text{normal } (0,\ EW_j^2)\ .$$

It follows immediately that

$$a_n(\hat{\beta}_n-\beta_0) = \frac{\frac{1}{a_n} \sum_{j=1}^{n} X_jW_j}{\frac{1}{a_n^2} \sum_{j=1}^{n} X_j^2\cdot} \Rightarrow \text{normal } (0,EW_j^2). \qquad \#$$

Based on the work by Resnick (1982) on point processes we can say a little more about the limiting distribution Z_1/Z_2 for case i). Let $\{E_j,W_j\}_{j=1}^{\infty}$ be an iid sequence of pairs of independent random variables where $E_j \sim$ exponential and $W_j \sim G$. Let $\Gamma_j = E_1+\ldots+E_j$. Then

$$(Z_1,Z_2) \overset{d}{=} \left(\sum_{j=1}^{\infty} \Gamma_j^{-1/\alpha} W_j,\ \ \sum_{j=1}^{\infty} \Gamma_j^{-2/\alpha} \right) .$$

REFERENCES

Blattberg, R. and T. Sargent (1971). Regression with Non-Gaussian Stable Disturbances: Some Sampling Results, Econometrica 39, 501-510.

Box, G. E. P. and G. M. Jenkins (1976). Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco.

Breiman, L. (1965). On Some Limit Theorems Similar to the Arc-sin Law, Th. Prob. and Appl. X, 323-331.

Brockwell, P. J. and R. A. Davis (1983). Recursive Prediction and Exact Likelihood Determination for Gaussian Processes, Tech. Rpt. 65, Dept. Statistics, Colorado State Univ., Fort Collins.

Cambanis, S. and G. Miller (1981). Linear Problems in pth Order and Stable Processes, Siam J. Appl. Math. 41, 43-69.

Cambanis, S. and A. R. Soltani (1982). Prediction of Stable Processes: Spectral and Moving Average Representations, Tech. Rpt. 11, Center for Stochastic Processes, Univ. No. Carolina, Chapel Hill.

deHaan, L. (1970). On Regular Variation and Its Application to the Weak Convergence of Sample Extremes, Math. Centre Tracts 32, Mathematisch Centrum, Amsterdam.

deHaan, L., E. Omey and S. I. Resnick (1982). Domains of Attraction and Regular Variation in R^d, Tech. Rpt. 8136/5, Econometric Institute, Erasmus Univ., Rotterdam.

Embrechts, P. and C. M. Goldie (1980). On Closure and Factorization Properties of Subexponential and Related Distributions, J. Austral. Math. Soc. (Series A) 29, 243-256.

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

Fuller, W. A. (1976). Introduction to Statistical Time Series, Wiley, New York.

Hampel, F. R. (1979). The Influence Curve and Its Role in Robust Estimation, J. Amer. Stat. Ass. 62, 1179-1186.

Hannon, E. J. and M. Kanter (1977). Autoregressive Processes with Infinite Variance, J. Appl. Prob. 14, 411-415.

Huber, P. J. (1964). Robust Estimation of a Location Parameter, Ann. Math. Stat. 35, 73-101.

Huber, P. J. (1981). Robust Statistics, Wiley, New York.

Kanter, M. and W. L. Steiger (1974). Regression and Autoregression with Infinite Variance, Adv. Appl. Prob. 6, 768-783.

Martin, R. D. and J. M. Jong (1977). Asymptotic Properties of Robust Generalized M-Estimates for the First-Order Autoregressive Parameter, Tech. Rpt., Bell Labs, Murray Hill.

Masreliez, C. J. and R. D. Martin (1977). Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter, IEEE Trans. Aut. Cont. AC-22, 361-371.

Resnick, S. I. (1982). Point Processes, Regular Variation and Weak Convergence, to appear in Adv. Appl. Prob.

Resnick, S. I. and P. Greenwood (1979). A Bivariate Stable Characterization and Domains of Attraction, J. Mult. Anal. 9, 206-221.

Rudin, W. (1974). Real and Complex Analysis, 2nd ed., McGraw-Hill, New York.

Sharpe, M. (1969). Operator Stable Distributions on Vector Spaces, Trans. Amer. Math. Soc. 136, 51-65.

Smith, V. K. (1973). Least Squares Regression with Cauchy Errors, Bull. Oxford Univ. Inst. of Econ. and Stat. 35, 223-231.

Stuck, B. W. (1978). Minimum Error Dispersion Linear Filtering of Scalar Symmetric Stable Processes, IEEE Trans. Aut. Cont. AC-23, 507-509.

Prepared By

APCOMP, Inc.
Lower Level, Suite 1
1311 South College
Fort Collins, Colorado 80524