DISCUSSION PAPERS IN ECONOMICS

Working Paper No. 00-06

Minimum Dispersion and Unbiasedness: 'Best' Linear Predictors for Stationary ARMA *a*-Stable Processes

Jonathan B. Hill
Department of Economics, University of Colorado at Boulder
Boulder, Colorado

September 2000

Center for Economic Analysis
Department of Economics

University of Colorado at Boulder
Boulder, Colorado 80309

Minimum Dispersion and Unbiasedness: "Best" Linear Predictors for Stationary ARMA α-Stable Processes

by

Jonathan B. Hill*

Dept. of Economics, University of Colorado at Boulder

Job Market Paper #1

The present paper is currently under review for publication in

Stochastic Processes and their Applications.

Abstract

Many volatile financial time series have been assumed to be driven by a distribution with an infinite population variance beginning with the seminal observations by Mandlebrot (1963) and Fama (1965). Although a flourish of statistical and econometric research

Key Words and Phrases: Best Linear Predictors, stable-law, infinite variance, covariation

AMS 1991 Subject Classifications. Primary: 60 G25; **Secondary:** 62 M10.

* *Dept. of Economics, Campus Box 256, University of Colorado at Boulder, Boulder Colo. 80309-0256*, jonathan.hill@colorado.edu

in the fields of estimation and inference for processes with infinite variance has been the result, few attempts have been made to characterize tractable and efficient prediction techniques. Existing techniques are restricted to the class of predictors which minimize a prediction error dispersion criterion. However, for linear optimal forecasts which are based on a finite sample of available data, minimum dispersion methods generally result in biased forecasts (i.e. $E[\hat{e}|\mathbf{X}] \neq 0$, with $\hat{e}$ the prediction error), or predictors which are not unique in the case when the population mean in infinite. Moreover, solution algorithms can be highly cumbersome. This analytical rift between the minimization of an appropriately chosen dispersion metric and canonical conditional expectations has largely been ignored in the literature.

In the present paper, we develop and completely characterize a theory of optimal linear forecasts for stationary $ARMA$ processes based on the criterion of unbiasedness (i.e. $E[\hat{e}|X_1, ..., X_n] = 0$). We demonstrate that the added prediction error dispersion becomes increasingly negligible as the dimension of X grows. Moreover, our method provides unique solutions even in the case when the population mean is infinite, is far less costly to compute, and obtains the minimum error dispersion of the class of linear predictors for any $AR(p)$ data generating process with $n \geq p$. Further, we develope a fast recursive solution algorithm comparable to traditional techniques that exist only for processes with finite variance. Finally, a numerical experiment demonstrates that the variability of the minimum dispersion prediction bias can be relatively extensive, while the prediction error dispersion of the best linear unbiased predictor is comparable to the minimum level of dispersion, suggesting that the simpler unbiased method may provide the "best" linear predictor.

1. Introduction

In the present paper, we are interested in methods of optimal linear forecasts of the causal-invertible $ARMA$ process $X_n - \phi_1 X_{n-1} - ... - \phi_p X_{n-p} = \epsilon_n + \theta_1 \epsilon_{n-1} + ... + \theta_q \epsilon_{n-q}$ where the innovations ϵ_t are *iid* non-normal α-stable. The stable-laws[1] have been the focus of a growing body of empirical and theoretical research in financial and macro-economics since the now classical studies by Mandlebrot (1965) and Fama (1963) on the distributional behavior of common asset prices. In particular, in those and subsequent studies (e.g. Cheng and Rachev, 1995, Jansen and de Vries, 1991, and McCulloch, 1984, 1987, 1996, 1997) the empirical distributional characteristics of asset returns, asset prices, option prices, and forward and spot exchange rates have in many cases displayed the characteristic of having "heavy distribution tails"; indeed, there exists evidence for a substantially high probability of large deviations. Importantly, the odds of large positive or negative swings are typically too great to be modelled by a normal distribution, and without further behavioral information (e.g. structural breaks[2]), choice of the assumed underlying

[1] Recall that any stable random variable with $\alpha \neq 1$ is represented by the log-c.f. $\ln \omega(\epsilon, t) = i\delta t - C|t|^{\alpha}(1 - i\beta[t/|t|]\tan[.5\pi\alpha])$, where $\delta \in R$, $C > 0$, $\beta \in [-1, 1]$ and $\alpha \in (0, 2)$ denote respectively the location, scale/dispersion, skewness and characteristic exponent. When $\beta = \delta = 0$, ϵ is symmetric α-stable ($S\alpha S$). It is well known that $\alpha = \sup(q : E|\epsilon|^q < \infty)$, implying $\forall \alpha < 2$ the population variance is infinite. In the multivariate case, $\alpha \neq 1$, for any jointly distributed stable $(n+1)$-vector $(\mathbf{X}, Y)$ we have

$$\ln \omega(\mathbf{t}, s) = i(\boldsymbol{\delta}, \mathbf{t}) - \int_{S_{n+1}} |(\mathbf{t}, \mathbf{s})|^{\alpha} [1 - i\mathrm{sgn}(s, t)\tan(.5\pi\alpha)]\, d\Gamma(\mathbf{t}, s)$$

where S_d is the unit hypersphere on R^d, and Γ denotes the stable spectral measure, a finite non-negative Borel measure on S_{n+1}. Special cases include $\alpha = 2$ (Gaussian), $\alpha = 1$ (Cauchy) and $\alpha = 1/2$ (inverted χ^2). See Ibragimov and Linnik (1971), Zoloratev (1971) and Samorodnitski and Taqqu (1994) for further exposition on stable processes. Throughout, we ignore the Cauchy case ($\alpha = 1$) for compactness, however all results can be extended to Cauchy r.v.'s.

[2] Competing literatures with the stable and stable-domain hypotheses exist in which volatile financial time-series are modeled as distrubutional "mixtures" of normal random variables, markov switching struc-

distribution would be necessarily arbitrary.

Stable random variables, however, have the property of infinite population variance (i.e. heavy tails) for all stable-laws except the normals, and are conveniently distributionally "stable" under addition: the sum of stables is itself a stable random variable. This property of stable summation is particularly attractive for modelling the behavioral properties of volatile low frequency (e.g. weekly/mothly) asset returns which are constructed as sums (products) of high frequency returns (e.g. daily/hourly) . For a general survey on the development of pricing models for stably-distributed assets, see McCulloch (1996). See, also, Gamrowski and Rachev (1999), and the citations therein, for the recent development of a stable-law *capital asset pricing model*.

Although research in statistical and econometric theory employing the stable-laws, or processes whose scaled sums converge to stable random variables, is extensive[3], surprisingly little exists in the econometric field of linear prediction and the theory of optimal forecasting. Our goal is to develop a simple and accurate projection technique for predict-

tures, or combinations of the above under a GARCH innovations structure. Analytical methods employed for testing the normal-mixtures hypothesis against the stable hypothesis, however, have recently come under scrutiny in McCulloch (1997). In this work, the nature of the null compound hypothesis is criticized (e.g. the *compound* null of *i.i.d.* stable is often rejected in favor of a normal-mixture), or, variously, empirical statistical methods for detection of processes governed by infinite variance stable-laws have been shown in simulation study to render biased rejections of the null in favor of alternative hyptheses.

[3]Adequately citing the existing literature is necessarily impossible, however without any attempt to be inclusive, we list a few papers and their topics here which reasonably represent the girth of topics covered in the field of estimation and inference with infinite variance processes: Hannan and Kanter (1977: *least squares*), Cline (1983: *least squares, M-estimation, prediction*), Phillips (1987: *unit root analysis*), Phillips and Loretan (1991: *Durbin-Watson test*), Knight (1993: *least absolute deviation, ARX estimation*), Kokoszka and Taqqu (1996: *moving average estimation with dependent errors*), Runde (1997: *tests for serial dependence*), Caner (1998: *cointegration tests*).

ing out-of-sample values of the infinite variance process X based on the available information $X_1, ..., X_n$. In particular, we develop methods which dramatically simplify existing solution mechanics for deriving the optimal linear forecast, and provide prediction which uniformly out-performs existing computationally burdensome techniques.

If we denote by L the lag operator, we have

$$\phi_p(L)X_t = \theta_q(L)\epsilon_t, \tag{1}$$

where we restrict the polynomials $\phi_p(L) = 1 + z\phi_1 + ... + z^p\phi_p$ and $\theta_p(L) = 1 + z\theta_1 + ... + z^q\theta_q$ to have no common roots, and to satisfy

$$\phi_p(L)\theta_p(L) \neq 0 \ \ \forall z \in R \text{ such that } |z| < 1. \tag{2}$$

In particular, we assume $\phi_p(L)$ has no roots outside the unit circle. It follows (see Brockwell and Davis, 1987, Cf. Cline, 1983) that there exist unique stationary solutions to (1),

$$X_n = \sum_{i=0}^{\infty} \pi_i \epsilon_{n-i} \qquad X_n = \epsilon_n + \sum_{i=1}^{\infty} \psi_i X_{n-i} \tag{3}$$

with $\sum_{i=1}^{\infty} |\pi_i|^{\delta} < \infty$ and $\sum_{i=1}^{\infty} |\psi_i|^{\delta} < \infty$, $\delta < \min(1, \alpha)$, and $|\pi_i| < \lambda\rho^i$ for some $\lambda > 0$ and $\rho \in (0, 1)$ as $i \to \infty$.

For a k-ahead forecast, we consider linear predictors of the truncated form

$$\hat{X}_{n+k} = \sum_{i=1}^{n} a_i X_{n+1-i} \tag{4}$$

$k \geq 1$. For non-normal stable-laws (i.e. $\alpha < 2$), the problem of developing an optimal linear projection is non-trivial, and produces an analytical rift that has largely been ignored in the literature. In particular, when $\alpha = 2$, the linear space of the Gaussian process is a Hilbert space which provides the theory of minimum mean squared error prediction criterion. This elegant theory has been fully developed in the Gaussian case:

the *mmse* predictor is optimally linear and identically the conditional expectations. The linear space of the stable-laws with $1 \leq \alpha < 2$, however, is a Banach space, and whenever $\alpha < 1$, only a metric space. Indeed, when $\alpha < 2$, in most cases we must choose between minimizing an acceptable L_ρ-metric, or employing the traditional conditional expectations form, $E\left[X_{n+k}|X_1, ..., X_n\right]$, or synonymously $E[X_{n+k} - \mathbf{a}'\mathbf{X}|X_{n+1-t}] = 0$. See Cambanis and Miller (1991: Theorems 5.6 and 5.8) for abstract examples which differentiate between conditional expectations and minimization of an appropriately chosen L_ρ-distance for $S\alpha S$ r.v.'s in stochastic integral form. Moreover, in many cases, the conditional expectations may not render a tractable linear form, or even exist, Cf. Hardin *et al* (1991) and Cambanis and Wu (1992). For example, provided ϵ_t belongs to the domain of attraction of a stable-law, few results exist which characterize the (linear) conditional expectations, Cf. Cioczek and Taqqu (1994). The central theme of this paper, therefore, is the complete analytical articulation of the unbiased linear prediction of stable linear processes based on conditional expectations and its performance when compared to traditional L_ρ-methods.

For processes with infinite variance, optimal linear predictors based on an L_ρ-criterion include time-domain minimum dispersion for univariate and multivariate processes which belong to the domain of attraction of a non-normal stable-law (Cline, 1983; Cline and Brockwell, 1985; and Soltani and Moeanaddin, 1994), minimization of the expected absolute prediction error and spectral-domain techniques (Cambanis and Soltani, 1982). In the univariate minimum dispersion case, a method widely cited in the literature, Cline and Brockwell (1985), Cf. Cline (1983), employ Stuck's (1978) criterion of minimizing the prediction error dispersion, with dispersion of a moving average defined as

$$disp(X_n) = disp\left(\sum_{i=1}^{\infty} \pi_i \epsilon_{n-i}\right) = \sum_{i=1}^{\infty} |\pi_i|^{\alpha}. \tag{5}$$

Whenever the *iid* variate ϵ_t is governed by a stable-law, $\alpha \in (0, 2)$, with scale $C_\epsilon > 0$, standard c.f. manipulation dictates that

$$X_n \stackrel{D}{=} \epsilon_1 \left(\sum_{i=1}^{\infty} |\pi_i|^{\alpha} \right)^{1/\alpha} \qquad C_{X_n}^{\alpha} = C_{\epsilon}^{\alpha} \sum_{i=1}^{\infty} |\pi_i|^{\alpha} < \infty \tag{6}$$

where the finiteness of the resulting scale follows necessarily from (3). Thus, minimization of the prediction error dispersion is identical to the problem of finding a vector **a** which minimizes the stable-scale associated with the prediction error, provided the error is stably distributed. In order to see that the prediction error $\hat{e} \equiv \hat{X}_{n+k} - \sum_{i=1}^{n} a_i X_{n+1-i}$ is governed by a stable-law $\forall n \geq 1$, observe that

$$\begin{aligned} X_{n+k} - \sum_{i=1}^{n} a_i X_{n+1-i} &= \sum_{i=0}^{\infty} \pi_i \epsilon_{n+k-i} - \sum_{j=1}^{n} a_j \sum_{i=0}^{\infty} \pi_i \epsilon_{n+1-j-i} \qquad (7) \\ &= \sum_{i=0}^{k-1} \pi_i \epsilon_{n+k-i} + \sum_{i=1}^{n} \left(\pi_{n+k-i} - a_1 \pi_{n-i} - ... - a_{n-i+1} \pi_0 \right) \epsilon_i \\ &\quad + \sum_{i=0}^{\infty} \left(\pi_{n+k+i} - a_1 \pi_{n+i} - ... - a_n \pi_i \right) \epsilon_{-i} \\ &= \sum_{i=0}^{\infty} \lambda_i \epsilon_{n+k-i} \stackrel{d}{=} \epsilon_1 \left(\sum_{i=0}^{\infty} |\lambda_i|^{\alpha} \right)^{1/\alpha}, \end{aligned}$$

where $\sum_{i=0}^{\infty} |\lambda_i|^{\alpha} < \infty$ follows from (3) and Cline (1983). Consequently, $disp(\hat{e}) = C_{\epsilon}^{\alpha} \sum_{i=0}^{\infty} |\lambda_i|^{\alpha}$, therefore without loss of generality we assume $C_{\epsilon} = 1$.

Moreover, if we define the signed-power $z^{<\delta>}$ in the usual manner (i.e. $z^{<\delta>} = |z|^{\delta} sgn(z)$), then from the FOC's of the minimization problem it is straightforward to show that $\forall t = 1..n$ and $\forall \alpha \in (0,2)$,

$$\begin{aligned} \frac{\partial \, disp(\hat{e})}{\partial a_t} &= - \sum_{i=k-1+t}^{n+k-1} \lambda_i^{<\alpha-1>} \pi_{i-(k-1+t)} - \sum_{i=n+k}^{\infty} \lambda_i^{<\alpha-1>} \pi_{i-(k-1+t)} \qquad (8) \\ &= - \sum_{i=0}^{n-t} \lambda_{i+k-1+t}^{<\alpha-1>} \pi_i - \sum_{i=n-t+1}^{\infty} \lambda_{i+k-1+t}^{<\alpha-1>} \pi_i \\ &= - \sum_{i=0}^{\infty} \lambda_{i+k-1+t}^{<\alpha-1>} \pi_i = 0, \quad t = 1..n. \end{aligned}$$

Thus, the set $\{\mathbf{a} : \mathbf{a} = \arg\min disp(X_{n+k} - \sum_{i=1}^{n} a_i X_{n+1-i})\}$ solves the homogenous system of n-equations implied by (8). Nonlinearity in the sequence $\{\lambda_i^{<\alpha-1>}\}_{i \geq 1}$ often implies a computationally extensive solution technique, and when $\alpha < 1$ the solution set is typically not unique. See Cline and Brockwell (1985: Theorem 3.2 and Lemma 4.1).

Importantly, the statistic $\sum_{i=0}^{\infty} \lambda_{i+k-1+t}^{<\alpha-1>} \pi_i$ is identically the stable *covariation* of X_{n+1-t} on the prediction error $\hat{X}_{n+k} - \sum_{i=1}^{n} a_i X_{n+1-i}$, denoted and defined as

$$\left[X_{n+1-t}, X_{n+k} - \sum_{i=1}^{n} a_i X_{n+1-i}\right]_\alpha \equiv \int_{S_{n+1}} u_t \left(\nu - \sum_{i=1}^{n} a_i u_{n+1-i}\right)^{<\alpha-1>} d\Gamma(\mathbf{u}, \nu), \quad (9)$$

where $(\nu, \mathbf{u})'$ denotes the vector $(X_{n+k}; X_1, ..., X_n)'$ in polar coordinates. Observe that for general α^{th} order processes, $\alpha < 2$, we use

$$\begin{aligned} &\left\langle X_{n+1-t}, X_{n+k} - \sum_{i=1}^{n} a_i X_{n+1-i} \right\rangle_\alpha \quad (10) \\ = \quad & E\left[X_{n+1-t}\left(X_{n+k} - \sum_{i=1}^{n} a_i X_{n+1-i}\right)^{<\alpha-1>}\right]. \end{aligned}$$

The stable covariation plays a central role in the development of our linear predictor, hence we will discuss several of this statistic's important properties. The interested reader is referred to Cambanis and Miller (1981) for pioneering work in the articulation of covariation and its conjugate for generalized α^{th} order processes, $\alpha < 2$. See, also, Samorodnitsky and Taqqu (1994) for a comprehensive treatment of stable dependence. Now, for any three stable-laws, say z_1, z_2, and z_3, $[z_1, z_2]_2 = cov(z_1, z_2)/2$. Moreover, z_1-independent-z_2 implies $[z_1, z_2]_\alpha = [z_2, z_1]_\alpha = 0$, however whenever $1 < \alpha < 2$ the obverse may not hold. Furthermore, additivity in the first argument is easy:

$$\begin{aligned} [az_1 + bz_2, z_3]_\alpha &= \int_{S_3} (au_1 + bu_2)\nu^{<\alpha-1>} d\Gamma(\mathbf{u}, \nu) \quad (11) \\ &= a\int_{S_3} u_1 \nu^{<\alpha-1>} d\Gamma(\mathbf{u}, \nu) + b\int_{S_3} bu_2 \nu^{<\alpha-1>} d\Gamma(\mathbf{u}, \nu) \\ &= a[z_1, z_3]_\alpha + b[z_2, z_3]_\alpha. \end{aligned}$$

Moreover, provided z_1 and z_2 are independent and symmetrically distributed (i.e. (z_1, z_2) ˜ $S\alpha S$), quasi-linearity in the second argument provides $[z_3, az_1 + bz_2]_\alpha = a^{<\alpha-1>}[z_3, z_1]_\alpha$

$+\ b^{<\alpha-1>}\left[z_3, bz_2\right]_\alpha$ by inspection of the stable characteristic function and spectrum, Cf. Miller (1977). The covariation, therefore, is not, in general, symmetric in its arguments (hence we say the "covariation of z_1 *on* z_2"). From the afore mentioned properties of additivity and independence, provided the sequence ϵ_t is *iid* it is easy to show that

$$[X_{n+1-t}, \hat{e}]_\alpha = \sum_{i=0}^{\infty} \lambda_{i+k-1+t}^{<\alpha-1>} \pi_i. \tag{12}$$

It follows immediately that the set $\{\mathbf{a} : [X_{n+1-t}, \hat{e}]_\alpha = 0,\ \ t = 1..n\}$ identically solves the minimum dispersion problem, *but not necessarily the conjugate set* $\{\mathbf{a} : [\hat{e}, X_{n+1-t}]_\alpha = 0,$ $t = 1..n\}$.

Precisely because of this lack of symmetry, in general, does the MDLP not imply the canonical condition of unbiasedness in the error term, $E[\hat{e}|\mathbf{X}] = 0$. Indeed, if $\alpha > 0$ and $E[\epsilon] = 0$, then $E[\hat{e}|\mathbf{X}] \neq 0$ implies the residuals are not *iid*, a fundamental condition by assumption. Further, the prediction bias itself can obtain substantial levels of dispersion for a non-negligible subset of the parameter space Θ, as will be discussed below. Moreover, because the best linear unbiased predictor is unique, all other prediction methods (e.g. Cambanis and Soltani, 1992; Soltani and Moeanaddin, 1994) render necessarily biased predictors. As a consequence, throughout the remainder of the paper for an appropriately chosen vector $\mathbf{a}$ we refer to the "alternative" system $[\hat{e}, X_{n+1-t}]_\alpha = \sum_{i=0}^{\infty} \lambda_{i+k-1+t} \pi_i^{<\alpha-1>}$ $= 0$ as *covariation-orthogonal* (i.e. $\boldsymbol{\lambda}_t(\mathbf{a})'\tilde{\boldsymbol{\pi}} = 0$ where we define the linear vector $\boldsymbol{\lambda}_t(\mathbf{a}) =$ $[\lambda_{0+k-1+t}(\mathbf{a}), ...,]'$, and $\tilde{\boldsymbol{\pi}} = [\pi_0^{<\alpha-1>}, ...])$, and we define the solution set $\{\mathbf{a} : [\hat{e}, X_{n+1-t}]_\alpha$ $= 0,\ \ t = 1..n\}$ as the *covariation-orthogonal linear predictor* (COLP) coefficients. In the sequel, we will demonstrate that the COLP is easier to compute, necessarily implies non-collinearity $E[\hat{e}|\mathbf{X}] = 0$, provides a unique solution set in all cases $\alpha \in (0, 2)$, and obtains the minimum level of dispersion for a wide class of linear processes. Due to solution uniqueness, it will be trivially the case that the COLP is the best predictor of the class of linear unbiased predictors.

The rest of the paper is organized as follows. In Section 2, we derive the relationship between the conditional expectations, and the unbiased and minimum dispersion linear predictors. We subsequently construct optimal projection functions for asymptotic and truncated predictors, and explicitly derive optimal solutions according to the covariation-orthogonality criterion for $AR(p)$ and $ARMA(1,1)$ processes in Sections 3 and 4. Section 5 follows with derivations for higher order moving average and $ARMA$ processes, including the characterization of a fast numeircal algorithm for solving the best linear unbiased predictor based on resursive prediction residuals. Section 6 concludes with a numerical comparison of the minimum dispersion and unbiased predictors.

2. Conditional Expectations and Covariation-Orthogonal Linear Predictors

Our investigation begins with an articulation of the relationship between multivariate conditional expectations, the COLP, and predictor unbiasedness when the innovations sequence ϵ_t is governed by an *iid* stable-law. Further details can be found in Cambanis and Wu (1992) and Samorodnitsky and Taqqu (1994). Throughout, we assume $\alpha \in (0,2)$, $\alpha \neq 1$.

Lemma 1 (Existence) *Let $X_1, ..., X_n; X_{n+k}$ be jointly α-stable with $\beta \in [-1,1]$, and let X_j satisfy (1) - (3). Then, provided*

$$\int_{S_{n+1}} (\nu - \mathbf{a}'\mathbf{u}) \left|\mathbf{t}'\mathbf{u}\right|^{\alpha-1} d\Gamma(\nu, \mathbf{u}) = 0 \tag{13}$$

for any $\mathbf{t} \in R^n$, the following criteria are necessarily and sufficiently identical, and subsequently render the same solution set $\mathbf{a}$:

i. $E[X_{n+k}|\mathbf{X}] = \mathbf{a}'\mathbf{X}$

ii. $E[X_{n+k} - \mathbf{a}'\mathbf{X}|\mathbf{X}] = 0$

iii. $\left[X_{n+k} - \mathbf{a}'\mathbf{X}, X_{n+1-t}\right]_\alpha = 0, \forall t = 1..n$

iv. $\left\langle X_{n+k} - \mathbf{a}'\mathbf{X}, X_{n+1-t}\right\rangle_\alpha = 0, \forall t = 1..n$

v. $\mathbf{a} = \arg\min_{\mathbf{a}\in R^n}\{disp(E\left[\hat{e}|X_1, ..., X_n\right])\}$.

Proof. Trivially, if there exists a vector $\mathbf{a} \in R^n$ such that (i) is true, then (ii) is immediate. Moreover, observe that whenever $E[X_{n+k} - \mathbf{a}'\mathbf{X}|\mathbf{X}] = \mathbf{0}$, then necessarily

$$\begin{aligned} \left\langle X_{n+1-t}, X_{n+k} - \mathbf{a}'\mathbf{X}\right\rangle_\alpha &= \\ E\left[\left(X_{n+k} - \mathbf{a}'\mathbf{X}\right) X_{n+1-t}^{<\alpha-1>}\right] &= E_x\left\{X_{n+1-t}^{<\alpha-1>} E\left[X_{n+k} - \mathbf{a}'\mathbf{X}|\mathbf{X}\right]\right\} \\ &= 0 \end{aligned} \tag{14}$$

where E_x denotes the expectation operator with respect to the σ-field generated by $\mathbf{X}$. Conversely, provided $\left\langle X_{n+k} - \mathbf{a}'\mathbf{X}, X_{n+1-t}\right\rangle_\alpha = 0$, and $E[X_{n+k}|\mathbf{X}] = \mathbf{b}'\mathbf{X}$ for some $\mathbf{b} \in R^n$ then $\forall t = 1..n$,

$$\sum_{i=1}^{n} (b_i - a_i)\, E\left[X_{n+1-t}^{<\alpha-1>} X_{n+1-i}\right] = 0$$

holds for *any* $ARMA(p, q)$ and for *any* $\alpha \in (0, 2)$ only if $b_i = a_i$, $i = 1..n$.

In order to prove that criteria $(i) - (iv)$ are identical, therefore, it suffices to prove that (i) exists and implies, necessarily and sufficiently, (iii). Now, by Theorems 1 and 2 of Cambanis and Wu (1992), $E[X_{n+k}|\mathbf{X}] = a'\mathbf{X}$ *a.s. if and only if* $\forall \mathbf{t} \in R^n$ and $\alpha \in (0, 2)$, $\alpha \neq 1$,

$$\begin{aligned} \int_{S_{n+1}} (\nu - \mathbf{a}'\mathbf{u})\,(\mathbf{t}'\mathbf{u})^{<\alpha-1>} d\Gamma(\nu, \mathbf{u}) &= 0 \\ \int_{S_{n+1}} (\nu - \mathbf{a}'\mathbf{u})\,|\mathbf{t}'\mathbf{u}|^{\alpha-1} d\Gamma(\nu, \mathbf{u}) &= 0. \end{aligned} \tag{15}$$

Therefore, simply chose n vectors $\mathbf{t}_i = [0, 0, ...0, 1, 0, ...0]'$ with the 1 in the i^{th} row. Consequently,

$$\int_{S_{n+1}} (\nu - \mathbf{a}'\mathbf{u})\,(u_i)^{<\alpha-1>} d\Gamma(\nu, \mathbf{u}) = 0 \quad i = 1..n \tag{16}$$

which is, by definition, (iii). We conclude this proof by establishing the identity between (i) and (v). Now, consider any linear predictor, say

$$\hat{X}_{n+k} = \sum_{i=1}^{n} b_i X_{n+1-i}, \tag{17}$$

and observe that we have the tautological regression form

$$X_{n+k} = \sum_{i=1}^{n} b_i X_{n+1-i} + \tilde{e} \tag{18}$$

where $\tilde{e} = X_{n+k} - \sum_{i=1}^{n} b_i X_{n+1-i}$ is governed by a stable-law $\forall b_i \in (-\infty, \infty)$, Cf. (7). Therefore, by (3) and (5), $disp(E\left[\tilde{e}|X_1, ..., X_n\right])$ is identically

$$\sum_{i=0}^{n-1} \left| \sum_{j=0}^{i} \pi_{i-j} \left(a_{j+1} - b_{j+1}\right) \right|^{\alpha} + \sum_{i=0}^{\infty} \left|\pi_{n+i} \left(a_1 - b_1\right) + ... + \pi_i \left(a_n - b_n\right)\right|^{\alpha},$$

which is trivially minimized by setting $b_j = a_j \; \forall j = 1..n$. ■

Remark 1: The identity of (i) - (v) holds for any (possibly maximally) skewed stable $n+1$ vector $(\mathbf{X}, Y)$. Indeed, in general we do not even require the stipulations detailed in (1) - (3). In order to explicitly derive the optimal coefficient set $\mathbf{a}$ under the proviso that (1) - (3) hold, however, we require that $(\mathbf{X}, Y)$ be jointly $S\alpha S$. We take this issue up in the sequel.

Remark 2: Consider the general unbiased predictor $E[Y|X_{1,}..X_n]$ where the vector $(X_1, .., X_n; Y)'$ is jointly distributed α-stable, and assume for the moment that the $X_i's$ are pair-wise mutually independent. Then $\int_{S_{n+1}} (\mathbf{a'u}) (u_i)^{<\alpha-1>} d\Gamma(\nu, \mathbf{u}) = a_i \int_S |u|^{\alpha} d\Gamma(u)$, $\forall i = 1..n$. Consequently,

$$a_i = \frac{[Y, X_i]_{\alpha}}{[X_i, X_i]_{\alpha}} \quad i = 1..n,$$

where $\mathbf{a} = \{a_j : [Y - \mathbf{a'X}, X_i]_{\alpha} = 0\}$ and $[X_i, X_i]_{\alpha} = \sum_{i=1}^{\infty} |\pi_i|^{\alpha}$. This trivial result, of course, holds only as benchmark for more sophisticated $ARMA$ processes.

Remark 3: The identity between criteria (i) and (iv) in the stable case allows for great simplification of solution methods. However, further non-stable processes which belong to the domain of attraction of a stable-law are not guaranteed to generate a linear conditional expectations. In particular, condition (*iii*) will not be appropriate, and conditions (*i*) and (*iv*) are certainly not necessarily implied. See Cioczek and Taqqu (1994).

We leave for Sections 3 and 4 details on the uniqueness of the solution set **a** in Lemma 1.

3. Linear Prediction with $AR(P)$ Processes

We begin by establishing a general result which characterizes the unique linear projection $\hat{X}_{n+k} = P(X_{n+k}, \mathbf{X})$ based on *covariation-orthogonality*. The following lemma establishes concretely that the COLP is the MDLP as $n \to \infty$ for any stationary *ARMA* α-stable processes. We subsequently derive the COLP for any finite order autoregressive process, and conclude the section by proving the COLP is the best linear unbiased estimator for any stationary *ARMA* α-stable processes and any $n \geq 1$. In order to exploit the property of quasi-linearity in the second argument of the covariation, we require the assumption that in all cases the vector $(X_1, ..., X_n; Y)$ ˜ $S\alpha S$. It should be pointed out that in the joint $S\alpha S$ case, (13) is trivially satisfied (Cambanis and Wu, 1992), hence it suffices to consider condition (*iii*) of Lemma 1.

Lemma 2 (Uniqueness) *For any* $ARMA(p, q)$ *process* X_n *assume (1)-(3) are true, and denote by* $\hat{S}$ *the class of random variables of the form*

$$\sum_{j=n+1}^{\infty} \rho_j \epsilon_j + \sum_{j=1}^{\infty} \nu_j X_{n+1-j} \tag{19}$$

where $\sum_{i=1}^{\infty} |\rho_j|^{\delta} < \infty$ *and* $\sum_{i=1}^{\infty} |\nu_j|^{\delta} < \infty$ *for some* $\delta < \min(1, \alpha)$*. Then, for any* $Y \in \hat{S}$*, the set*

$$P_n^Y = \left\{ \sum_{i=1}^{n} a_i X_{n+1-i} : E\left[Y - \sum\nolimits_{i=1}^{n} a_i X_{n+1-i} | X_n, ..., X_1 \right] = 0 \right\}$$

consists of exactly one element, $\hat{Y}_n = \sum_{i=1}^{n} \nu_i X_{n+1-i}$, as $n \to \infty$. Moreover, the mapping $Y \to \hat{Y}_n$ is linear on $\hat{S}$. Further, as $n \to \infty$ the random variable $\hat{Y}_n$ is the unique element of the set

$$\acute{P}_n^Y = \left\{ \sum_{i=1}^{n} a_i X_{n+1-i} : disp\left(Y - \sum_{i=1}^{n} a_i X_{n+1-i} \right) \text{ is minimized} \right\}.$$

Proof. For any $Y \in \hat{S}$,

$$\begin{aligned} Y - \sum_{i=1}^{n} a_i X_{n+1-i} &= \sum_{j=n+1}^{\infty} \rho_j \epsilon_j + \sum_{j=1}^{n} \epsilon_{n+1-j} \left[\sum_{i=1}^{j} (\nu_i - a_i)\, \pi_{j-i} \right] \\ &+ \sum_{j=n+1}^{\infty} \epsilon_{n+1-j} \left[\sum_{i=1}^{n} (\nu_j - a_j)\, \pi_{j-i} + \sum_{i=n+1}^{j} \nu_j \pi_{j-i} \right], \end{aligned} \tag{20}$$

consequently, by identities (ii) and (iii) of Lemma 1, $\forall t = 1..n$, it suffices to solve

$$\left[Y - \sum_{i=1}^{n} a_i X_{n+1-i}, X_{n+1-t} \right]_\alpha = \sum_{j=0}^{n-t} \lambda_{j+t} \pi_j^{<\alpha-1>} + \sum_{j=n+1}^{\infty} \varphi_{j+t} \pi_j^{<\alpha-1>} = 0 \tag{21}$$

where

$$\lambda_j = \sum_{i=1}^{j} (\nu_i - a_i)\, \pi_{j-i} \quad \varphi_j = \sum_{i=1}^{n} (\nu_i - a_i)\, \pi_{j-i} + \sum_{i=n+1}^{j} \nu_i \pi_{j-i}. \tag{22}$$

In the limit, the sufficiency of $a_j = \nu_j$, $j = 1..n$, is trivial. Now, by (21) for $t = n$, $\lim_{n\to\infty} \left(\sum_{i=1}^{n} (\nu_i - a_i)\, \pi_{n-i} \right) \pi_0^{<\alpha-1>} = 0$, and when $t = n - 1$,

$$\lim_{n\to\infty} \left(\sum_{i=1}^{n-1} (\nu_i - a_i)\, \pi_{n-1-i} \right) \pi_0^{<\alpha-1>} + \left(\sum_{i=1}^{n} (\nu_i - a_i)\, \pi_{n-i} \right) \pi_1^{<\alpha-1>} = 0.$$

For non-degenerate sequences $\{\pi_j\}_{j\geq1}$, we conclude necessarily that $\nu_n = a_n$ as $n \to \infty$. It follows recursively that as $n \to \infty$ equality in (21) holds *if and only if* $a_j = \nu_j$, $j = 1..n$, hence the unique element of the set P_n^Y as $n \to \infty$ is $\hat{Y}_n = \sum_{i=1}^{n} \nu_i X_{n+1-i}$, as claimed.

The linearity of the mapping $Y \to \hat{Y}_n$ follows easily by observing that for any $Y = \sum_{i=1}^{m} b_i Y_i$, $Y_i \in \hat{S}$, $i = 1..n$, then

$$Y = \sum_{j=n+1}^{\infty} \tilde{\rho}_j \epsilon_j + \sum_{j=1}^{\infty} \tilde{\nu}_j X_{n+1-j},$$

where $\tilde{\rho}_j = \sum_{i=1}^n b_i \rho_{j,i}$ and $\tilde{\nu}_j = \sum_{i=1}^n b_i \nu_{j,i}$. Therefore $Y \in \hat{S}$, and $a_j = \tilde{\nu}_j$, $j = 1, 2, ...$ Consequently,

$$\begin{aligned} \acute{P}_n^Y(Y) &= \hat{Y} = \sum_{j=1}^{\infty} \tilde{\nu}_j X_{n+1-j} = \sum_{j=1}^{\infty} \sum_{i=1}^{n} b_i \nu_{j,i} X_{n+1-j} \\ &= \sum_{i=1}^{n} b_i \sum_{j=1}^{\infty} \nu_{j,i} X_{n+1-j} = \sum_{i=1}^{n} b_i \hat{Y}_i. \end{aligned}$$

Finally, as a consequence of Cline and Brockwell (1985: Lemma 2.1), Cf. (23), below, $\hat{Y}_n$ is the unique element of the set $\acute{P}_n^Y$. ■

Remark 1: The set P_n^Y defines the unique covariation orthogonal projection $\hat{S} \rightarrow \overline{span}(X_1, X_2, ...)$. In Corollary 3, below, we augment this result to any truncated projection.

Remark 2: Subsequent to (8), $\hat{Y}_n$ satisfies $\lim_{n\rightarrow\infty}[X_{n+1-t}, Y - \sum_{i=1}^n a_i X_{n+1-i}]_\alpha$ $= 0$ $\forall t = 1, 2, ..$ Consider, then, the covariation of X_{n+1-t} on the prediction error for any $t = 1..n$,

$$\begin{aligned} &\left[X_{n+1-t}, Y - \sum_{i=1}^{n} a_i X_{n+1-i} \right]_\alpha \qquad (23) \\ = &\sum_{j=0}^{n-1} \pi_j \left[\sum_{i=1}^{j+t} (\upsilon_i - a_i) \pi_i \right]^{<\alpha-1>} + \sum_{j=n+1}^{\infty} \pi_j \left[\sum_{i=1}^{n} (\upsilon_i - a_i) \pi_{j-i} + \sum_{i=n+1}^{j} \upsilon_i \pi_{j-i} \right]^{<\alpha-1>}. \end{aligned}$$

For any finite sample size n, the MDLP solution set $\mathbf{a}$ will not in general render the solution to (21) $\forall \alpha \in (0, 2)$ due to the non-linear components. However, as $n \rightarrow \infty$, a recursive argument identical to the exposition below (22) proves that the COLP minimizes the prediction error dispersion. In the sequel, we explicitly characterize the nature of the difference between the truncated COLP and MDLP methods for $ARMA(1,1)$ and $MA(q)$ processes.

Observe that lemmas 1 and 2 trivially imply the COLP is the unique best predictor of the class of linear unbiased predictors $\forall n \geq 1$.

Corollary 3 *For any ARMA(p, q) process X_n assume (1)-(3) are true, and denote by $\hat{S}$ the class of random variables defined in Lemma 2. Moreover, denote by E_0 the event that $E[Y - \sum_{i=1}^{n} a_i X_{n+1-i} | X_1, ..., X_n] = 0$. Consequently, for any $Y \in \hat{S}$, the set*

$$\tilde{P}_n^Y = \left\{ \sum_{i=1}^{n} a_i X_{n+1-i} : disp \left(Y - \sum_{i=1}^{n} a_i X_{n+1-i} | E_0 \right) \text{ is minimized} \right\}$$

consists of exactly one element, $\hat{Y}_n = \sum_{i=1}^{n} \tilde{\nu}_i X_{n+1-i}$, for any $n \geq 1$ where $\tilde{\nu}_i$ solves the following implicit system of n-equations,

$$\sum_{i=1}^{n} a_i \lambda_{i,t} = \sum_{i=1}^{n} \lambda_{i,t} \tilde{\nu}_{i,t}, \tag{24}$$

where

$$\begin{aligned} \lambda_{i,t} &= \pi_i \sum_{j=1-t}^{n-t} \pi_j^{<\alpha-1>} + \sum_{j=n+1}^{\infty} \pi_j^{<\alpha-1>} \pi_{j-i} \qquad (25) \\ \tilde{\nu}_{i,t} &= \nu_i + \frac{\omega_{n,t}}{n\lambda_{i,t}} \\ \omega_{n,t} &= \sum_{i=0}^{t} \nu_{n+1+i} \left(\sum_{j=0}^{\infty} \pi_{j+t-1} \pi_{n+1+j}^{<\alpha-1>} \right) + \sum_{i=t+1}^{\infty} \nu_{n+1+i} \left(\sum_{j=0}^{\infty} \pi_j \pi_{n+1+j}^{<\alpha-1>} \right). \end{aligned}$$

Moreover, the mapping $Y \to \hat{Y}_n$ is linear on $\hat{S}$. Finally, $\lim_{n\to\infty} \tilde{\nu}_i = \lim_{n\to\infty} \tilde{\nu}_{i,t} = \nu_i$ for every $i = 1, 2, ...$

Remark 1: The set $\tilde{P}_n^Y$ defines the unique truncated covariation orthogonal projection of $\hat{S} \to \overline{span}(X_1, ...X_n)$, the space of all linear combinations of the available data.

For large $n < \infty$, (21) and (22) demonstrate that the truncated COLP will be approximately the MDLP as n grows large. Conversely, for large n the MDLP will be roughly unbiased. However, for general $AR(p)$ processes, the following result establishes that the truncated COLP is identically the MDLP provided $n \geq p$.

Theorem 4 *Let $X_1, ..., X_n; X_{n+k}$ be jointly α-stable with $\alpha \in (0, 2)$, and assume (1) - (3) holds with $q = 0$, $n \geq p$. Then, provided $k = 1$, (i)- (v) of Lemma 1 each provide a_j*

$= \phi_j$, $j = 1..p$. *Moreover,* $\forall k \geq 1$ *we obtain the recursive relationship*

$$\hat{X}_{n+k} = \phi_1 \hat{X}_{n+k-1} + ... + \phi_p \hat{X}_{n+k-p} \tag{26}$$

where optimally $\hat{X}_j = X_j \ \forall j \leq n$. *Further,* $\hat{X}_{n+k}$ *obtains the minimum level of error dispersion.*

Proof. (26) is immediate by Lemma 2 and Corollary 3. For the resulting minimized level of dispersion, see Cline and Bockwell (1985: Lemma 2.3 and Corollary 2.4) for derivation of the MDLP $\hat{X}_{n+k}$. ■

Remark 1: The best linear unbiased predictor $\hat{X}_{n+k}$ is precisely the least squares predictor for finite autoregressive processes. This elegant symmetry, however, does not hold for general $ARMA(p,q)$ processes, a topic we discuss in the sequel.

4. Linear Prediction of the $ARMA(1,1)$ α-Stable Process In the present section, we explicitly treat the random variable X_n which is governed by a stationary $ARMA(1,1)$ processes

$$X_n - \phi_1 X_{n-1} = \epsilon_n + \theta_1 \epsilon_{n-1} \tag{27}$$

with $|\phi| < 1$ and $|\theta| < 1$. Further, we demonstrate that the $AR(1)$ optimal COLP is identically the MDLP (which also follows from Theorem 3), however for truncated predictors, the MDLP is in general biased for higher order $ARMA$ processes. Observe that the following results hold for any characteristic exponent $\alpha \in (0,2)$, $\alpha \neq 1$.

Theorem 5 *Let* $(X_1, ..., X_n; X_{n+k})$ *be jointly* $S\alpha S$, *and assume (3) holds. Then (i) - (v) of Lemma 1 each provide the unique solution set*

$$a_j^k = (-\theta)^{j-1} \phi^{k-1} \left[\frac{(\theta + \phi)(1 - \eta + \xi) - \xi \eta^{n-j} (\eta \phi + \theta)}{1 - \eta + \xi(1 - \eta^n)} \right] \quad j = 1..n, \tag{28}$$

where we define

$$\xi = \frac{|\theta+\phi|^\alpha}{1-|\phi|^\alpha} \quad \eta = \theta\left[(\theta+\phi)^{<\alpha-1>} - \phi^{k-1}\right]. \tag{29}$$

The corresponding prediction error is

$$\begin{aligned} &1+\xi\left(1-|\phi|^{\alpha(k-1)}\right)+ \\ &|\theta^n|^\alpha \left|\frac{\phi^{k-1}\xi(1-\eta)}{1-\eta+\xi(1-\eta^n)}\right|^\alpha \left[\sum_{j=0}^{n-1}|\tilde{\eta}^j|^\alpha + 1\right], \end{aligned}$$

where $\tilde{\eta} = (\theta+\phi)^{<\alpha-1>} - \phi^{k-1}$. *Moreover, as* $n \to \infty$, *the set of coefficients* $\{a_j^k\}_{j\geq 1}$ *minimizes the prediction error.*

Proof. Because $|\phi| < 1$, we can write

$$X_t = \epsilon_t + (\theta+\phi)\sum_{j=1}^{\infty}\phi^{j-1}\epsilon_{t-j} \tag{30}$$

therefore, observe that $X_{n+k} - \mathbf{a}'\mathbf{X}$ reduces to

$$\begin{aligned} &\epsilon_{n+k} + \sum_{j=1}^{k-1}(\theta+\phi)\,\phi^{k-j-1} - \sum_{j=k}^{n+k-1}\left[a_{j-k+1} + (\theta+\phi)\sum_{i=0}^{j-k}a_i\phi^{j-k-i}\right]\epsilon_{n+k-j} \quad (31) \\ &-(\theta+\phi)\left[\sum_{i=0}^{n}a_i\phi^{n-i}\right]\sum_{j=n+k}^{\infty}\phi^{j-n-k}\epsilon_{n+k-j}. \end{aligned}$$

Now, define the sequence $\{c_j\}_{j=0}^n$ according to the recursive relationship $a_j^k = c_j - \phi c_{j-1}$with $c_0 = -\phi^{k-1}$ (i.e. $c_j = \sum_{i=0}^{j}\phi^i a_{j-i}^k$). Consequently, by the properties of covariation and the assumption that the sequence $\{\epsilon_j\}_{j\geq 1}$ is *iid* α-stable, after some manipulation it can be shown that $\forall\alpha \in (0,2)$ and $\forall t = 1..n-2$

$$\begin{aligned} &\left[X_{n+k} - \sum_{i=1}^{n}a_iX_{n+1-i}, X_{n+1-t}\right]_\alpha \quad (32) \\ =\ &\theta c_{t-1} + \left[1+(\theta+\phi)^{<\alpha-1>}\theta\right]c_t + (\theta+\phi)^{<\alpha-1>}\left(1+\theta\phi^{<\alpha-1>}\right)\sum_{j=0}^{n-t-2}\phi^j c_{t+1+j} \\ &+\left(\phi^{n-t-1}\right)^{<\alpha-1>}\left[(\theta+\phi)^{<\alpha-1>} + \phi^{<\alpha-1>}\left(\frac{|\theta+\phi|^\alpha}{1-|\phi|^\alpha}\right)\right]c_n \\ =\ &0, \end{aligned}$$

for $t = n-1$

$$
\begin{aligned}
\left[X_{n+k} - \sum_{i=1}^{n} a_i X_{n+1-i}, X_{n+1-t}\right]_\alpha &= \theta c_{n-2} + \left[1 + (\theta+\phi)^{<\alpha-1>}\theta\right] c_{n-1} \\
&\quad + \left[(\theta+\phi)^{<\alpha-1>} + \phi^{<\alpha-1>}\left(\frac{|\theta+\phi|^\alpha}{1-|\phi|^\alpha}\right)\right] c_n \\
&= 0,
\end{aligned} \tag{33}
$$

and when $t = n$,

$$
\left[X_{n+k} - \sum_{i=1}^{n} a_i X_{n+1-i}, X_{n+1-t}\right]_\alpha = \theta c_{n-1} + \left[1 + \frac{|\theta+\phi|^\alpha}{1-|\phi|^\alpha}\right] c_n = 0. \tag{34}
$$

This system of n-equations offers a recursive solution in the following stepwise manner. From (34),

$$
c_n = -\left(\frac{\theta}{1+\xi}\right) c_{n-1}, \tag{35}
$$

which is substituted into (33) providing

$$
c_{n-1} = -\theta c_{n-2}\left[\frac{1+\xi}{1+\xi\,(1+\eta)}\right]. \tag{36}
$$

Subsequently, employing (32) recursively, we deduce that

$$
c_j = -\theta c_{j-1}\left[\frac{1+\xi\sum_{i=1}^{n-j}\eta^{i-1}}{1+\xi\sum_{i=1}^{n}\eta^{i-1}}\right]. \tag{37}
$$

Observing that $c_0 = -\phi^{k-1}$, we obtain

$$
c_j = -\phi^{k-1}(-\theta)^j\left[\frac{1+\xi\sum_{i=1}^{n-j}\eta^{i-1}}{1+\xi\sum_{i=1}^{n}\eta^{i-1}}\right] = -\phi^{k-1}(-\theta)^j\left[\frac{1-\eta+\xi(1-\eta^{n-j})}{1-\eta+\xi(1-\eta^{n})}\right] \tag{38}
$$

provided $|\eta| < 1, \forall(\theta,\phi) \in \Theta$, which follows readily, and the optimal vector $\mathbf{a}$ can be found by solving $a_j^k = c_j - \phi c_{j-1}$. The proof is complete upon observing that

$$
\begin{aligned}
& disp(X_{n+k} - \mathbf{a}'X) \\
= \;& 1 + |\theta+\phi|^\alpha\left(\frac{1-|\phi|^{\alpha(k-1)}}{1-|\phi|^\alpha}\right) + \sum_{j=1}^{n}|c_j + \theta c_{j-1}|^\alpha + \left(\frac{|\theta+\phi|}{1-|\phi|^\alpha}\right)|c_n|^\alpha \\
= \;& 1 + \xi\left(1-|\phi|^{\alpha(k-1)}\right) + \sum_{j=1}^{n}|c_j + \theta c_{j-1}|^\alpha + \xi|c_n|^\alpha.
\end{aligned}
$$

Subsequently, from (38) we deduce

$$\begin{aligned}
c_j + \theta c_{j-1} &= \phi^{k-1}(-\theta)^j \left(\frac{\xi(1-\eta)\eta^{n-j}}{1-\eta+\xi(1-\eta^n)} \right) \\
|c_j + \theta c_{j-1}| &= |\theta^n|^\alpha \left| \phi^{k-1}\theta^{j-n}\eta^{n-j} \left(\frac{\xi(1-\eta)}{1-\eta+\xi(1-\eta^n)} \right) \right|^\alpha \\
&= |\theta^n|^\alpha |\tilde{\eta}^j|^\alpha \left| \frac{\phi^{k-1}\xi(1-\eta)}{1-\eta+\xi(1-\eta^n)} \right|^\alpha ,
\end{aligned}$$

hence

$$\begin{aligned}
& disp(X_{n+k} - \mathbf{a}'X) \\
= \; & 1 + \xi\left(1 - |\phi|^{\alpha(k-1)}\right) + |\theta^n|^\alpha \left| \frac{\phi^{k-1}\xi(1-\eta)}{1-\eta+\xi(1-\eta^n)} \right|^\alpha \left[\sum_{j=0}^{n-1} |\tilde{\eta}^j|^\alpha + 1 \right] \\
\to \; & 1 + \xi\left(1 - |\phi|^{\alpha(k-1)}\right)
\end{aligned}$$

as $n \to \infty$, the minimum level of dispersion $\forall\alpha > 0$ and $\forall k \geq 1$. ■

Remark 1: In all cases $\alpha \in (0, 2)$ the projection $X_{n+k} \to \hat{X}_{n+k}$ in unique and linear on $\overline{span}(X_1, X_2, ...)$. Comparatively, $\forall\alpha < 1$ the MDLP is not unique and depends on known parameter values. Further, by Corollary 3, the solution set (28) renders the best predictor of the class of linear unbiased predictors.

Remark 2: The set of COLP coefficients is remarkably similar to the MDLP coefficients for characteristic exponents $1 < \alpha < 2$. Indeed, if we redefine $\eta = |\theta|^{1/\alpha-1}$ and $\xi = [(|\theta + \phi|^\alpha)/(1 - |\phi|^\alpha)]^{1/\alpha-1}$, the coefficient set denoted in (28) will minimize the prediction error dispersion, Cf. Cline and Brockwell (1985). Moreover, clearly $\lim_{n\to\infty} a_j^k = \phi^{k-1}(-\theta)^{j-1}(\theta + \phi)$ which is identically the asymptotic MDLP coefficient set, Cf. Cline and Brockwell (1985: eq. (3.3)), thus demonstrating Lemma 2.

Remark 3: Theorem 5 immediately implies a pseudo-linearity property of the truncated COLP. Specifically, $\forall\alpha \in (0, 2)$ and for any $Y = \sum_{j=0}^{k-1} b_{j+1}X_{n+k-j}$, we obtain the best linear unbiased predictor $\hat{Y} = \sum_{j=0}^{k-1} b_{j+1}\hat{X}_{n+k-j}$ where $\hat{X}_{n+j} = \sum_{i=1}^{n} a_i^j X_{n+1-i}$

follows from (28). Moreover,

$$\hat{Y} = \left(b_1 + b_2\phi + ... + b_k\phi^{k-1}\right)\hat{X}_{n+1},$$

which follows by recursively solving (32) - (34) with $c_0 = -\left(b_1 + b_2\phi + ... + b_k\phi^{k-1}\right)$. This linear structure is identical to the truncated MDLP linearity property except, of course, for the unbiased one step-ahead predictor $\hat{X}_{n+1}$ itself. It follows trivially that any linear combination of biased minimum dispersion linear predictors will be biased.

Remark 4: By the definition of covariation, when $\alpha = 2$, the COLP is identically the least squares predictor. See Brockwell and Davis (1983) and Cline and Brockwell (1985) for accounts of optimal recursive methods for deriving linear predictors in the finite variance case..

The following corollary is an immediate result of Theorems 4 and 5.

Corollary 6 *Let $(X_1, ..., X_n; X_{n+k})$ be jointly $S\alpha S$, and assume (3) holds. Provided X_n is $ARMA(1,0)$ and $|\phi| < 1$, $a_1 = \phi^k$ and $a_i = 0\ \forall i > 1$. In particular, the COLP minimizes the prediction error dispersion. Moreover, whenever X_n is $ARMA(0,1)$ with $k = 1$,*

$$a_j = -(\theta^{j-1})\left[\frac{1 - |\theta|^{\alpha(n+1-j)}}{1 - |\theta|^{\alpha(n+1)}}\right] \quad j = 1..n. \tag{39}$$

Remark 1: Observe that in the $AR(1)$ case, the optimal k-ahead COLP is merely $\hat{X}_{n+k} = \phi^k X_n$, and in the $MA(1)$ for any $k \geq 2$, $\hat{X}_{n+k} = 0$.

Example We consider two $ARMA(1,1)$ cases. Let $X_t = .3X_{t-1} + .8\epsilon_{t-1} + \epsilon_t$, and put $C_\epsilon = 1$, $\alpha = 1.75$, $n = 3$ and $k = 1$. As a consequence of Theorem 5 and Cline and Brockwell (1985: Theorem 3.2) the COLP and MDLP are respectively

$$\begin{aligned}\hat{X}_4^C &= .9922X_3 - .6164X_2 + .2542X_1 \\ \hat{X}_4^M &= .8959X_3 - .5435X_2 + .2339X_1.\end{aligned}$$

The resulting levels of prediction error dispersion and $disp(E[\hat{e}|X_1,...,X_n])$ are

$$\begin{aligned} disp(\hat{e}^C) &= .16252 & disp(E\left[\hat{e}^C|X_1,...,X_n\right]) = 0.0 \\ disp(\hat{e}^M) &= .15046 & disp(E\left[\hat{e}^M|X_1,...,X_n\right]) = .70313. \end{aligned}$$

Finally, let $X_t = .9X_{t-1} - .25\epsilon_{t-1} + \epsilon_t$, and put $C_\epsilon = 1$, $\alpha = 1.2$, $n = 5$ and $k = 5$. Then,

$$\begin{aligned} \hat{X}_{10}^C &= .42740X_5 + .10625X_4 + .026414X_3 + .0065739X_2 + .0021326X_1 \\ \hat{X}_{10}^M &= .42647X_5 + .10662X_4 + .026654X_3 + .0066641X_2 + .0023058X_1, \end{aligned}$$

and

$$\begin{aligned} disp(\hat{e}^C) &= 1.9936 & disp(E\left[\hat{e}^C|X_1,...,X_n\right]) = 0.00 \\ disp(\hat{e}^M) &= 1.9932 & disp(E\left[\hat{e}^M|X_1,...,X_n\right]) = .038501. \end{aligned}$$

The COLP compares well in each case, providing a comparatively low level of prediction error dispersion for larger values of n and k. Indeed, the optimal coefficients are nearly identical in the second case due to the relatively large persistence parameter $\phi = .9$. Further, in both cases $disp(E\left[\hat{e}^M|X_1,...,X_n\right]) - disp(E\left[\hat{e}^C|X_1,...,X_n\right]) > disp(\hat{e}^C) - disp(\hat{e}^M)$, suggesting the magnitude of bias in the MDLP "outweighs" its comparative improvement in prediction error dispersion. We verify this claim in the Section 6 for any $n \in [1, 100]$. Moreover, the MDLP renders the symmetric result of diminishing the level of bias as k and n increase.

5. Optimal COLP for $MA(q)$ and $ARMA(p,q)$ α-Stable Processes Consider the finite $MA(q)$ processes X_n,

$$X_n = \theta_1\epsilon_{n-1} + ... + \theta_q\epsilon_{n-q} + \epsilon_n, \tag{40}$$

where $\theta_q(L) \neq 0$. It is easy to show that $\hat{X}_{n+k} = 0$ for any $k > q$, hence fix $k \in [1, q]$. As before, assume $(X_1, ..., X_n; X_{n+k})$ is jointly $S\alpha S$ with $\alpha \in (0, 2)$ $\alpha \neq 1$.

Theorem 7 *For any $\alpha \in (0, 2)$, $\alpha \neq 1$, the optimal COLP solution set $\mathbf{a}$ satisfies $a_i = \det\left(M_i^0\right) / \det(M)$, $i = 1..n$, where M denotes an $(n + 2q) \times (n + 2q)$ identity matrix with the sub-region $(q + 1, 1) : (n+q, n+2q)$ defined by the following $n \times (n + 2q)$ matrix*

$$\begin{bmatrix} \xi_q & \xi_{q-1} & \xi_{q-2} & \cdots & \xi_{1-n-q} \\ \xi_{q+1} & \xi_q & \xi_{q-1} & \cdots & \xi_{2-n-q} \\ . & . & . & \cdots & . \\ \xi_{q+n-1} & \xi_{q+n-2} & \xi_{q+n-3} & \cdots & \xi_{-q} \end{bmatrix} \tag{41}$$

where we define $\xi_j = \sum_{i=0}^{q} \theta_i^{<\alpha-1>} \theta_{i+j}$, and M_i^0 denotes the matrix M with the i^{th} column replaced by the $(n + 2q)$-vector

$$\begin{bmatrix} 0 & \ldots & 0 & \xi_k & \ldots & \xi_{n+k-1} & 0 & \ldots & 0 \end{bmatrix}',$$

where the first and last sub-vectors of zeros are of q-length. The solution set $\{a_j\}_{1-q}^{n+q}$ is such that $a_j = 0$, $j \leq 0$, $j > n$, and $\theta_j = 0$, $j < 0$, $j > q$, $\theta_0 = 1$.

Proof. Observe that

$$\begin{aligned} & X_{n+k} - \sum_{i=1}^{n} a_i X_{n+1-i} \\ = \; & \sum_{j=0}^{k-1} \epsilon_{n+k-j}\theta_j - \sum_{j=1}^{n+q} \epsilon_{n+1-j}\left(a_{j-q}\theta_q + \ldots + a_{j-1}\theta_1 + a_j - \theta_{k+j-1}\right) \end{aligned} \tag{42}$$

where $a_j = 0$, $j < 0$, $j > n$, and $a_0 = -1$. Consequently, after some manipulation, we have $\forall t = 1..n$,

$$\begin{aligned} \left[X_{n+k} - \sum_{i=1}^{n} a_i X_{n+1-i}, X_{n+1-t}\right]_\alpha &= \\ -\sum_{j=t-q}^{q+t} a_j \sum_{i=0}^{q} \theta_i^{<\alpha-1>}\theta_{-j+t+i} + \sum_{j=0}^{q} \theta_j^{<\alpha-1>}\theta_{j+k+t-1} &= 0 \end{aligned} \tag{43}$$

where $\theta_j = 0$, $j < 0$, $j > n$, $\theta_0 = 1$ and $a_j = 0$, $j \leq 0$, $j > n$. The above implicit linear system of equations can be immediately solved upon application of Crámer's Rule, which completes the proof. ■

Remark 1: The solution set $\mathbf{a}$ is unique $\forall \alpha \in (0, 2)$. Observe that $\forall \alpha < 1$ the MDLP reduces the choice set of $\mathbf{a}$ to $\binom{n+q}{q}$ possibilities as functions of the solutions to the polynomial $z^q + \theta_1 z^{q-1} + ... + \theta^q = 0$, Cf. Cline and Brockwell (1985: p. 293). Thus, the computational burden will be comparatively extensive for higher order moving average processes for large n.

Finally, consider the general $ARMA(p, q)$ processed denoted in (1) - (3), and define the following sequence

$$\varphi_1 = 1, \quad \varphi_i = \sum_{j=1}^{i} \psi_{j-i+1} \varphi_{i-1} \quad i = 2..k, \tag{44}$$

where the coefficients $\{\psi_j\}_{j \geq 1}$ are defined in (3). Then, by (3), for any $k \geq 1$ we deduce

$$X_{n+k} = \sum_{j=0}^{k-1} \rho_j \epsilon_{n+k-j} + \sum_{j=1}^{\infty} \nu_j X_{n+1-j} \tag{45}$$

where $\rho_j = \sum_{i=1}^{k} \varphi_i$, $j = 1..k$ and $\nu_j = \sum_{i=1}^{k} \psi_{j+k-i} \varphi_i$, $j = 1, 2, ...$ Consequent to Lemma 2, $X_{n+k} \in P_n^Y$ and $a_j = \nu_j$, $j = 1..n$, is the unique solution as $n \to \infty$. Moreover, from (44) and (45) it is straightforward to verify the following result.

Corollary 8 *For any $ARMA(p, q)$ process X_n such that (1)-(3) hold, and for any $k \geq 1$ there exists a unique covariation orthogonal linear predictor which minimizes the prediction error as $n \to \infty$. Specifically, this linear predictor satisfies the recursive relationship*

$$\hat{X}_{n+k} = \sum_{j=1}^{k-1} \psi_j \hat{X}_{n+k-j} + \sum_{j=k}^{\infty} \psi_j X_{n+k-j}. \tag{46}$$

In the case of ARMA(p, q) prediction when only a finite sample is available, we offer the subsequent general corollary to Lemma 2 for truncated $ARMA(p, q)$ predictors.

Corollary 9 *For any ARMA(p, q) process X_n such that (1)-(3) hold, for any $k \geq 1$, $n \leq \infty$ and $\alpha \in (0, 2)$, $\alpha \neq 1$, there exists a unique COLP coefficient set* $\mathbf{a}$ *such that (i) - (v) of Lemma 1 are satisfied. In particular, a_j solves*

$$\sum_{i=1}^{n} a_i \lambda_{i,t} = \sum_{i=1}^{n} \nu_i \lambda_{i,t} + \sum_{i=0}^{t} \nu_{n+1+i} \left(\sum_{j=0}^{\infty} \pi_{j+t-1} \pi_{n+1+j}^{<\alpha-1>} \right) + \sum_{i=t+1}^{\infty} \nu_{n+1+i} \left(\sum_{j=0}^{\infty} \pi_j \pi_{n+1+j}^{<\alpha-1>} \right), \tag{47}$$

where $\forall t = 1..n$,

$$\lambda_{i,t} = \pi_i \sum_{j=1-t}^{n-t} \pi_j^{<\alpha-1>} + \sum_{j=n+1}^{\infty} \pi_j^{<\alpha-1>} \pi_{j-i} \tag{48}$$

$$\nu_j = \sum_{i=1}^{k} \psi_{j+k-i} \varphi_i$$

and $\pi_j = 0$, $j < 0$.

Proof. Subsequent to remark 1 of Lemma 2, $\forall t = 1..n$, $[X_{n+k} - \mathbf{a}'\mathbf{X} | X_{n+1-t}]_\alpha = 0$ implies (47), and (48) is immediate due to (44) and (45). The unique coefficient set $\mathbf{a}$ follows from traditional methods for solving linear systems of equations. ■

Remark 1: We may infer from (47) and (48) that for large n, the truncated predictor $X_{n+k}^* = \sum_{j=1}^{k-1} \psi_j X_{n+k-j}^* + \sum_{j=k}^{n+k-1} \psi_j X_{n+k-j}$ will be close to optimal. Of course, conversely the truncated MDLP will be approximately unbiased for large n.

Solving the implied n-equation system in (47) will be difficult even for low order processses. However, existing algorithmic techniques that employ recursive prediction residuals can be easily extended to the stable laws. For the following derivations, we require some compact notation, in addition to stable-representations detailed in previous sections. Define the recursive prediction residual $e_k = X_{k+1} - \hat{X}_{k+1}$, $k = 0...n-1$, where $\hat{X}_1 = 0$ by convention. Additionally, for each horizon $h \in \mathfrak{N}$, define the real-valued sequence $\{\theta_{n,i}^{(h)}\}_{i=1}^{n}$

such that a best linear unbiased predictor, cf. Lemma 1 and Corollary 9, satisfies the recurisive residuals formula, $\hat{X}_{n+h} = \sum_{i=1}^{n} \theta_{n,i}^{(h)}(X_{n+1-i} - \hat{X}_{n+1-i})$. The following theorem delivers a recursive algorithm for one-step ahead stable prediction of univariate processes by utilizing a well-known prediction residuals format that has been characterized only in the case of Hilbert-space prediction; see, e.g., Brockwell and Davis (1987).

Theorem 10 *For any stable-law X_t, the unbiased α-orthogonal one-step ahead predictor $E[X_{n+1}|X_n, ..., X_1] = \hat{X}_{n+1}$ satisfies $\hat{X}_{n+1} = \sum_{i=1}^{n} \theta_{n,i}^{(1)}(X_{n+1-i} - \hat{X}_{n+1-i})$, where for each $k = 0...n-1$,*

$$\theta_{n,n-k}^{(1)} = \left([X_{n+1}, X_{k+1}]_\alpha - \sum_{i=0}^{k-1} \theta_{n,n-i}^{(1)} [e_i, X_{k+1}]_\alpha \right) [e_k, X_{k+1}]_\alpha^{-1}. \tag{49}$$

Proof. Consider the recursive prediction residual form, $\hat{X}_{n+1} = \sum_{i=1}^{n} \theta_{n,n-i}^{(1)}(X_{n+1-i} - \hat{X}_{n+1-i})$, for any $n > 0$. Differencing with respect to X_{n+1} and applying the α-orthogonality condition, cf. Lemma 1, to both sides of the equality with respect to an arbitrary element X_{k+1}, $k \in [0, n-1]$, we obtain

$$\left[X_{n+1} - \hat{X}_{n+1}, X_{k+1}\right]_\alpha = \left[X_{n+1} - \sum_{i=1}^{n} \theta_{n,i}^{(1)} e_{n-i},\ X_{k+1}\right]_\alpha. \tag{50}$$

Now, observe that by α-orthogonality, the unbiased linear predictor necessarily renders $[e_j, e_k]_\alpha = 0$ for any $j > k$. Consequently, $[X_{n+1} - \hat{X}_{n+1}, X_{k+1}]_\alpha = 0$, therefore the principles of covariation (linearity in the first argument) and (50) imply

$$\begin{aligned} 0 &= [X_{n+1}, X_{k+1}] - \sum_{i=0}^{k} \theta_{n,n-i}^{(1)} [e_i, X_{k+1}]_\alpha \qquad (51) \\ &= [X_{n+1}, X_{k+1}] - \sum_{i=0}^{k-1} \theta_{n,n-i}^{(1)} [e_i, X_{k+1}]_\alpha - \theta_{n,i}^{(1)} [e_k, X_{k+1}]_\alpha. \end{aligned}$$

Solving, we obtain

$$\theta_{n,n-k}^{(1)} = \left([X_{n+1}, X_{k+1}]_\alpha - \sum_{i=0}^{k-1} \theta_{n,n-i}^{(1)} [e_i, X_{k+1}]_\alpha \right) [e_k, X_{k+1}]_\alpha^{-1}, \tag{52}$$

which proves the result. ■

Remark 1: The substantive distinction between between the classical "residuals" Hilbert-space algorithm and the above stable-law procedure lies in the non-symmetry of the α-orthogonality condition, $[e_i, e_j]_\alpha$. Indeed, by construction, $[X_{i+1} - \hat{X}_{i+1}, X_{i+1-j}]_\alpha$ $= 0$ for every $j = 1, 2, ...i$, however, as detailed in Section 1, the covariation is in general not symmetric in it arguments. Thus, while $[e_i, e_j]_\alpha = 0$, $i > j$, may be true, $[e_j, e_i]_\alpha$ $= 0$ does not typically follow. A simple counter-example exists for the case when the prediction residuals are independent, which is in general not true in the present setting, by construction. Of course, when $\alpha = 2$, the L_2-inner product is symmetric and (49) reduces to the classical formula. See Brockwell and Davis (1987).

6. Numerical Experiment In the preceding sections, we proved that in theory the COLP the MDLP will be "similar" for large n. It is left for experimentation to demonstrate the nature of the differential between the two predictors. Therefore, in this concluding section, we explicitly derive that $MA(1)$ and $ARMA(1,1)$ COLP and MDLP coefficients, denoted respectively $a_{c,j}$ and $a_{m,j}$, over a 99 × 99 parameter grid $(\alpha, \theta) \subseteq [1.01, 1.99] \times [.01, .99]$, and for various autoregressive parameters $\phi \in \{0, .7\}$. Results for negative parameter values are symmetrically identical, and recall that the COLP and MDLP are identical for all finite order autoregressive processes. Additionally, it is easy to show that step values $k \to \infty$ monotonically diminishes both measures of dispersion while rendering results comparatively similar to one-step ahead prediction, , hence we report experiments only for one step-ahead forecasts (i.e. $k = 1$). We explicitly ignore processes with low order characteristic exponents (i.e. $\alpha < 1$) in order to exploit the uniqueness of the MDLP for purposes of criterion comparability. For each linear projection problem and arbitrary $n \in \{1, 10, 50, 100\}$ we plot the differential

$$disp(X_{n+k} - \mathbf{a}'_m\mathbf{X}) - disp(X_{n+k} - \mathbf{a}'_c\mathbf{X}) \tag{53}$$

over the computed grid, and report the extrema

$$\begin{aligned} &\min_{(\alpha,\theta)\subseteq[1.01,1.99]\times[.01,.99]} \{disp(X_{n+k} - \mathbf{a}'_m\mathbf{X}) - disp(X_{n+k} - \mathbf{a}'_c\mathbf{X})\} \\ &\max_{(\alpha,\theta)\subseteq[1.01,1.99]\times[.01,.99]} \{disp(X_{n+k} - \mathbf{a}'_m\mathbf{X}) - disp(X_{n+k} - \mathbf{a}'_c\mathbf{X})\}. \end{aligned} \tag{54}$$

Additionally, in order to demonstrate the bias in the canonical expected prediction error of the MDLP, we plot the dispersion of the conditional expectations of the MDLP prediction error,

$$\tilde{d}_{MDLP} = \sum_{j=1}^{n} |z_j + \theta z_{j-1}|^{\alpha} + \left(\frac{|\theta + \phi|^{\alpha}}{1 - |\phi|^{\alpha}}\right) z_n, \tag{55}$$

where $z_j = \sum_{i=0}^{j} \phi^{j-i}[(a_{c,i} - \phi a_{c,i-1}) - (a_{m,i} - \phi a_{m,i-1})]$. Of course, by Corollary 4 the COLP is unbiased, hence $\tilde{d}_{COLP} = 0$, thus $\tilde{d}_{MDLP}$ also serves as the differential. Finally, for each $n \in [1, 100]$ and $\phi \in \{0, .7\}$ we calculate the number of grid points for which

$$\begin{aligned} |disp(X_{n+k} - \mathbf{a}'_m\mathbf{X}) - disp(X_{n+k} - \mathbf{a}'_c\mathbf{X})| &< 10^{-8} \\ \left|\tilde{d}_{MDLP} - \tilde{d}_{COLP}\right| &< 10^{-8} \end{aligned} \tag{56}$$

and subsequently plot the frequency progressions.

In summary, clearly, the MDLP provides an improvement over the COLP based on a criterion of prediction error dispersion over a non-negligible sub-set of the parameter space for small sample sizes (see Figures 1 and 2). However, there exists a nearly one-to-one correspondence between the regions in which the COLP prediction error dispersion is large and the prediction error of the MDLP based on conditional expectations is large (see Figures 3 and 4). Indeed, and not surprisingly, the greater the improvement in prediction "stability" by employment of the MDLP, *ceteris paribus*, the greater the resulting prediction error bias, a problem classically associated with efficient, biased predictors/estimators.

In the $MA(1)$ case, Figure 5 demonstrates that there exists a substantial monotonic improvement over the dispersion differential as a function of n. As n grows large, the parameter region over which the differential is negligible (i.e. less than 10^{-8}) increases monotonically. However, as $\phi \nearrow 1$ in general $ARMA(1,1)$ models, the dispersion differential is affected non-linearly. Indeed, the relative spatial improvement of both differentials is nearly identical in the $MA(1)$ case. However, for large $|\phi|$ values, the MDLP improves on bias faster than the COLP improves on prediction error dispersion when n is small. For large n, however, clearly the unbiased COLP improves on prediction error dispersion at a greater rate.

For instance, by Theorem 4, for small moving average components $|\theta|$, we expect the COLP to be close to the MDLP, as evinced by the above example. Therefore, for $ARMA(1,1)$ processes with weak inter-temporal persistence the MDLP seems to be the preferred projection method *based on a criterion of minimum error dispersion*. As the number of data points grows large and for large parameter values $|\phi|$, based on the same criterion the COLP compares well with the best linear predictor. However, all evidence suggests that the variability of the prediction error bias of the MDLP outweighs the benefits associated with minimization of the prediction error dispersion.

We argue, in conclusion, that the unbiased linear forecast is the "best" linear predictor (indeed, it is the best predictor of the class of linear *unbiased* predictors). The unbiased predictor renders a combined predictor error dispersion and error bias that is uniformly lower than that provided by minimum dispersion methods. Morevoer, and perhaps more importantly from a computational point of view, due to a reduction in the non-linearity of the criterion function, the unbiased predictor is universally easier to compute than existing methods. The result is a dominant method for deriving optimal linear forecasts of highly volatile time-series.

References

Brockwell, P.J. and R.A. Davis, 1983, Recursive Prediction and Exact Likelihood Determination for Gaussian Processes, Technical Report 65, Dept. of Statistics, Colorado State University.

Brockwell, P.J. and R.A. Davis, 1987, Time Series: Theory and Methods (Springer: New York).

Cambanis, S. and I. Fakhre-Zakeri, On Prediction of Heavy-Tailed Autoregressive Sequences: Forward Versus Reverse Time, Theory and Probability Applications 39, 217-233.

Cambanis, S. and G. Miller, 1981, Linear Problems in p^{th} Order and Stable Processes, SIAM Journal of Mathematics 41, 43-69.

Cambanis, S and A.R. Soltani, 1982, Prediction of Stable Processes: Spectral and Moving Average Representations, Tech. Report 11, Center for Stochastic Processes (Univ. of N. Carolina, Chapel Hill).

Cambanis, S. and W. Wu, 1992, Multiple Regression on Stable Vectors, Journal of Multivariate Analysis 41, 243-272.

Caner, Mehmet, 1998, Tests for Cointegration with Infinite Variance Errors, Journal of Econometrics 86, 155-175.

Cheng, B.N. and S.T. Rachev, 1995, Multivariate Stable Futures Prices, Mathematical Finance 5, 133-153.

Cioczek, R. and M. Taqqu, 1994, Does Asymptotic Linearity of the Regression Extend to Stable Domains of Attraction? Journal of Multivariate Analysis 48, 70-86.

Cline, D.B.H., 1983, Estimation and Linear Prediction for Regression, Autoregression and ARMA with Infinite Variance Data, Unpublished Ph.D. Dissertation (Department of Statistics, Colorado State University).

Cline, D.B.H. and P. Brockwell, 1985, Linear Prediction and ARMA Processes with Infinite Variance, Stochastic Processes and Their Applications 19, 281-296.

Fama, E., 1965, Portfolio Analysis in a Stable Paretian Market, Management Sciences 11, 404-419.

Gamrowski, B. and S.T. Rachev, 1999, A TEstable Version of the Pareto-Stable CAPM, Mathematical and Computer Modelling 29, 61-81.

Hannan, E.J. and Marek Kanter, 1977, Autoregressive Processes with Infinite Variance, Journal of Applied Probability 14, 411-415.

Hardin, Clyde, Gennady Samorodnitsky and Murad Taqqu, 1991, Nonlinear Regression of Stable Random Variables, The Annals of Applied Probability 1, 582-612.

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

Jansen, Dennis and Casper de Vries, 1991, On the Frequency of Large Stock Returns: Putting Booms and Busts into Perspective, The Review of Economics and Statistics 30, 18-24.

Knight, K., 1993, Estimation in Dynamic Linear Regression Models with Infinite Variance Errors, Econometric Theory 9, 570-588.

Kokoszka, Piotr and M.S. Taqqu, 1996, Infinite Variance Stable Moving Averages with Long Memory, Journal of Econometrics 73, 79-99.

Mandlebrot, B., 1963, The Variation of Certain Specualtive Prices, Journal of Business 36, 394-419.

McCulloch, J.H., 1984, Stable Option Prices, Ohio State University, Dept. of Economics.

McCulloch, J.H., Foreign Exchange Option Pricing with Log-Stable Uncertiantly. In S.J. Khoury and A. Ghosh, edc., *Recent Developments in Internation Banking and Finance* 1, 231-245.

McCulloch, J.H., 1996, Financial Applications of Stable Distributions, Handbook of Statistics vol. 14 (Elsevier: New York).

McCulloch, J. Huston, 1997, Measuring Tail Thickness to Estimate the stable Index : A Critique, Journal of Business and Statistics 15, 74-81.

Miller, G., 1977, Some Results on Symmetric Stable Distributions and Processes, Institute of Statistics no. 1121 (Univ. of N. Carolina, Chapel Hill).

Phillips, P.C.B., 1990, Time Series Regression with a Unit Root and Infinite-Variance Errors, Econometric Theory 6, 44-62.

Phillips, P.C.B. and Mico Loretan, 1991, The Durbin-Watson Ratio under Infinite Variance Errors, Journal of Econometrics 47, 84-114.

Runde, Ralf, 1997, The Asymptotic Null Distribution of the Box-Pierce Q-statistic for Random Variable with Infinite Variance: An Application to German Stock Returns, Journal of Econometrics 78, 205-216.

Samorodnitsky, G. and M.S. Taqqu, 1989, Conditional Moments of Stable Random Variables, Stochastic Processes.

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

Soltani, A. and R. Moeanaddin, 1994, On Dispersion of Stable Random Variables and its Applications in the Prediction of Multivariate Stable Processes, Journal of Applied Probability 31, 691-699.

Stuck, B.W., 1978, Minimum Dispersion Linear Filtering of Scalar Symmetric Stable Processes, IEEE Trans. Aut. Cont. 23, 507 - 509.

Zolotarev, V.M., 1971, One Dimension Stable Distributions (American Mathematical Society, Providence).