Fitting Autoregressive Models via Yule-Walker Equations Allowing Heavy Tail Innovations

Robert Serfling[1]

University of Texas at Dallas

July 2010

[1]Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. Email: `serfling@utdallas.edu`. Website: `www.utdallas.edu/~serfling`. Support by a Society of Actuaries Grant, NSF Grant DMS-0805786 and NSA Grant H98230-08-1-0106 is gratefully acknowledged.

Abstract

Modern treatments of actuarial risk decision problems increasingly involve heavy tailed data and distributions. Here we consider the setting of time series and the problem of fitting an autogressive model with heavy tailed innovations. Assuming only finite first moments, we introduce a linear system of equations similar to the least squares approach but using Gini covariances instead of the usual ones. This leads to convenient, easily interpreted, closed form expressions for the estimated parameters, thereby capturing the advantages of the least squares approach without requiring second order moment assumptions. A "Gini autocovariance function" is introduced, along with certain other novel types. Related results for the multiple regression problem with heavy tailed errors are treated briefly as well. A numerical comparison of the Gini approach and the least squares approach is provided.

AMS 2000 Subject Classification: Primary 62M10 Secondary 62G05

Key words and phrases: Autoregression; Heavy tails; Gini covariance; Yule-Walker equations; Multiple regression

1 Introduction

Time series arise in a wealth of contexts in science, engineering, industry, and planning, including actuarial science (see Frees, 2011). Here the regression approach is very appealing for its simplicity and interpretability, as in other statistical settings, and takes the form of autoregressive (AR) models, which are especially attractive and useful. For a discrete-time formulation, the AR(p) model (e.g., Brockwell and Davis, 1991, and Frees, 2011) has the general form

$$X_t = \phi_0 + \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t, \ t = 0, \pm 1, \pm 2, \ldots, \qquad (1)$$

although an actual data set involves only a finite sequence of "times" $\{X_1, \ldots, X_T\}$. The goal is to determine $\phi_0, \phi_1, \ldots, \phi_p$, which are possibly complex functions of the underlying distributions of the variables X_t and ε_t. Assuming existence of the relevant means, along with $E(\varepsilon_t) = 0$, all t, we may obtain ϕ_0 via

$$\phi_0 = E(X_t) - \sum_{j=1}^{p} \phi_i E(X_{t-j}), \qquad (2)$$

so that it suffices to discuss how to obtain $\phi_1, \ldots, \phi_p$. It is standard to assume *finite second order moments* for the variables and that $\mathrm{Cov}(X_s, \varepsilon_t) = 0$ for $s < t$. Then taking certain covariances across (1), one obtains a convenient linear system of equations, the *Yule-Walker equations*, for the parameters $\phi_1, \ldots, \phi_p$ (see Brockwell and Davis, §8.2). A key role is played by the autocovariance function, whose definition imposes finiteness of $E(\varepsilon_t^2)$ for all t. Similarly, Yule-Walker *estimates* of $\phi_1, \ldots, \phi_p$ are obtained as explicit closed-form expressions in terms of sample covariances, facilitating straightforward computation, easy interpretation, and goodness-of-fit diagnostics.

It is becoming increasingly important, however, to allow *heavy tailed* distributions and data, for which second order moment assumptions are too stringent. In actuarial science, such data arise in loss modeling (Klugman, Panjer, and Wilmot, 2004), nursing home financing (Rosenberg, Frees, *et al.*, 2007, and Frees, 2011, Ch. 17), and hospital utilization and cost data (Rosenberg and Farrell, 2008), to mention a few settings involving either multiple regression or autoregression. We ask, therefore: *Can Yule-Walker estimates of $\phi_1, \ldots, \phi_p$ be obtained without imposing second order moment assumptions?*

Here an affirmative answer to this question is developed, making use of *Gini covariances* and a *Gini autocovariance function*, defined under just *first order moment assumptions*. Besides $E(\varepsilon_t) = 0$ as mentioned above, we assume that for each t the "innovation" ε_t is *independent* of the variables $X_s, s < t$, and that the time series $\{X_t\}$ is *strictly stationary* in the usual sense (e.g., Brockwell and Davis, 1991), with F denoting the common distribution of the variables X_t. For broad applicability, we make neither higher order moment nor parametric distributional assumptions, and thus the modeling setting (1) accommodates both *heavy tailed data* and *nonparametric applications*.

There already exist approaches requiring only first order (or even no) moment assumptions. The least absolute deviations (LAD) method (e.g., Bloomfield and Steiger, 1983, and Ling, 2005) requires essentially only a first order moment assumption, but it does not yield yield closed form expressions or simply solve a linear system. Another approach introduces an analogue of the usual autocorrelation function having a sample version that estimates it consistently as the sample length increases (see Davis and Resnick, 1985a, 1985b, 1986, and Brockwell and Davis, 1991, §13.3), but this "autocorrelation" function lacks straightforward interpretation, however. Resnick (1997) reviews this asymptotic approach and its limitations, and Feigin and Resnick (1999) discuss broadly the pitfalls of attempting to fit an autoregressive model with heavy-tailed innovations. Our "Gini" approach, on the other hand, possesses the computational features of least squares while requiring only first order moment assumptions. Even in the case of finite variances, our estimators are less sensitive to extreme observations in the data than those based on the usual sample covariances.

Section 2 presents background on Gini covariance, introduces a Gini autocorrelation function for the time series context, and treats sample estimates and their properties. In Section 3, a general covariance-type method of deriving Yule-Walker equations for the AR problem is introduced and applied using Gini covariances and autocorrelation. Detailed treatment of the "Gini AR" approach, including numerical illustrations with data, is provided in Section 4.

Closely related in structure to the AR model is the multiple regression (MR) model, which postulates a linear relationship between a designated "response" variable Z and selected "explanatory" variables (or "covariates" or "regressors") $Y_1, \ldots, Y_p$, i.e.,

$$Z = \phi_0 + \sum_{i=1}^{p} \phi_i Y_i + \varepsilon. \tag{3}$$

As with the AR model, if the relevant means exist and the "error" ε has *finite mean* $E(\varepsilon) = 0$, then ϕ_0 may be obtained via

$$\phi_0 = E(Z) - \sum_{i=1}^{p} \phi_i E(Y_i), \tag{4}$$

and it suffices to discuss how to obtain $\phi_1, \ldots, \phi_p$. Again, the usual second order moment assumptions preclude heavy tailed data. However, a "Gini" approach similar to that for the AR model, assuming that ε is *independent* of the variables $Y_1, \ldots, Y_p$, yields a linear system analogous to the usual "normal" equations produced by the least squares approach. In fact, a "Gini" approach to the MR problem already exists (Olkin and Yitzhaki, 1992, and Schechtman, Yitzhaki, and Pudalov, 2010), but our approach provides a novel derivation and yields a clarified view of the "Gini MR" approach and its properties, facilitating computation and applications. Brief discussion is provided in Section 5.

2 Gini covariance and autocovariance

The usual covariance and autocovariance extend the notion of variance to two variables considered together. An alternative to the variance for measuring spread is the Gini mean difference, and analogous extensions of it to two variables considered together are provided by the "Gini" covariance and autocovariance.

2.1 Definitions and Properties

Gini mean difference

For a *univariate* distribution F, an important alternative to the usual standard deviation as a measure of *spread* was introduced by Gini (1912):

$$\alpha(F) = E|X_1 - X_2| = E(X_{2:2} - X_{1:2}), \tag{5}$$

where X_1 and X_2 are independent observations having distribution F and $X_{1:2} \leq X_{2:2}$ denote their ordered values. It is now known as the *Gini mean difference* (GMD). We note that $\alpha(F)$ is finite if F has *finite mean.* For X having distribution F, let $\alpha(X)$ also denote $\alpha(F)$. See Yitshaki and Olkin (1991) for general discussion of the GMD and Serfling and Xiao (2007) for treatment within the broader context of *L-moments.*

A useful equivalent expression for $\alpha(F)$ is the *L-functional* (weighted integral of quantiles) representation

$$\alpha(F) = 2\int_0^1 F^{-1}(u)\,(2u-1)\,du, \tag{6}$$

where $F^{-1}(u) = \inf\{x : F(x) \geq u\}$, $0 < u < 1$, denotes the usual quantile function of F. This enables estimation of $\alpha(F)$ by substitution of a sample version of F. Another productive representation is

$$\alpha(X) = 2\mathrm{Cov}(X, 2F(X) - 1) = 4\,\mathrm{Cov}(X, F(X)), \tag{7}$$

which facilitates an illuminating interpretation: $\alpha(X)$ is 4 times the covariance of X and its *"rank"* in the distribution F. Both (6) and (7) yield

$$\alpha(X) = 2\int x\,(2F(x) - 1)\,dF(x), \tag{8}$$

still another useful expression.

Gini covariance

For a *bivariate* random vector (X, Y) with joint distribution F and respective marginal distributions F_X and F_Y having finite means, two alternatives to the usual covariance for measuring *dependence* of X and Y are the *Gini covariance of X with respect to Y*,

$$\beta^{(X,Y)}(F) = 2\mathrm{Cov}(X, 2F_Y(Y) - 1) = 4\,\mathrm{Cov}(X, F_Y(Y)), \tag{9}$$

and the *Gini covariance of* Y *with respect to* X, $\beta^{(Y,X)}(F)$. For convenience we denote these simply by $\beta(X,Y)$ and $\beta(Y,X)$.

Note that $\beta(X,Y)$ is proportional to the covariance of X and the F_Y-rank of Y, while $\beta(Y,X)$ is proportional to the covariance of Y and the F_X-rank of X. Thus we need not have equality of $\beta(Y,X)$ and $\beta(X,Y)$, the two together providing complementary pieces of information about the dependence of X and Y. Of course, if the variables X and Y are "interchangeable", i.e., (X,Y) and (Y,X) identically distributed, then $\beta(Y,X)$ and $\beta(X,Y)$ are equal. If X and Y are independent, then $\beta(Y,X) = \beta(X,Y) = 0$. Also, $\beta(X,X)$ is just $\alpha(X)$.

We note that the definition (9) parallels the representation (7) for $\alpha(X)$. Also, paralleling (8), we have

$$\beta(X,Y) = 2\int\int x\,(2F_Y(y) - 1)\,dF(x,y). \tag{10}$$

Further, paralleling (5), we have

$$\beta(X,Y) = E(X_{[2:2]} - X_{[1:2]}), \tag{11}$$

where (X_1,Y_1) and (X_2,Y_2) are independent observations on F, and, for $i = 1,2$, $X_{[i:2]}$ denotes the X-value or "*concomitant*" matched with $Y_{i:2}$, with $Y_{1:2} \leq Y_{2:2}$ the ordered values of Y_1 and Y_2 (see David and Nagaraja, 2003).

The notion of Gini covariance dates back at least to Daniels (1944), but see Schechtman and Yitzhaki (1987) and Yitshaki and Olkin (1991) for detailed treatment and perspective, including discussion of scaling of Gini covariance to produce a scale-free *Gini correlation* (of X with respect to Y),

$$\widetilde{\rho}(X,Y) = \frac{\beta(X,Y)}{\alpha(X)} = \frac{\mathrm{Cov}(X, F_Y(Y))}{\mathrm{Cov}(X, F_X(X))}. \tag{12}$$

This compares with the usual Pearson correlation

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y} \tag{13}$$

defined under second order moment assumptions.

For treatment of Gini covariance and correlation within the context of welfare economics and finance, and including certain extended versions, see Schechtman, Yitzhaki, and Artsev (2008). For treatment within the general context of *L-comoments* and including "L-coskewness" and "L-cokurtosis", see Serfling and Xiao (2007). The latter includes rigorous proof of the properties

(I) $-1 \leq \widetilde{\rho}(X,Y) \leq 1$

(II) $\widetilde{\rho}(X,Y) = \pm 1$ is attained if and only if X is a monotone function of Y and

(III) $\widetilde{\rho}(X,Y) = \rho(X,Y)$ when

a) both are defined,

b) F_X and F_Y are affinely equivalent in distribution, i.e., $F_Y(y) = F_X(\eta^{-1}(y-\theta)$ for some θ and η, and

c) X has linear regression on Y, i.e., $E(X \mid Y) = a + bY$ for some a and b.

Property (I) is true also for $\rho(X,Y)$, but property (II) is not.

For a *d-variate* random vector $(X_1, \ldots, X_d)$, we form the $d \times d$ *Gini covariance matrix* $[\beta(X_i, X_j)]_{d\times d}$ from all the pairwise Gini covariances. This analogue of the usual covariance matrix of pairwise covariances is defined under merely first moment assumptions.

Gini autocovariance function

In treating the AR(p) model (1), we will be interested in the Gini covariances of lag k, i.e., for observations k time units apart. These are $\beta(X_{1+k}, X_1)$ and $\beta(X_1, X_{1+k})$, for $k = 0, \pm 1, \ldots$. It is now convenient to adopt the assumption that the time series $\{X_t\}$ is *strictly stationary*. Let F denote the common distribution of the variables $\{X_t\}$.

By the stationarity assumption, $\beta(X_s, X_t) = \beta(X_{s+r}, X_{t+r})$ for any $r = 0, \pm 1, \ldots$. In particular, $\beta(X_{1+k}, X_1) = \beta(X_1, X_{1-k})$, for each k. Hence the two series of Gini covariances $\{\beta(X_{1+k}, X_1),\ k = 0, \pm 1, \pm 2, \ldots\}$ and $\{\beta(X_1, X_{1+k}),\ k = 0, \pm 1, \pm 2, \ldots\}$ are identical sets of numbers (arranged in different order) and thus provide the same information about the dependence structure of $\{X_t\}$. Therefore, adopting just one of these series of covariances, we define the *Gini autocovariance function* as

$$\gamma^{(\mathrm{G})}(k) = \beta(X_{1+k}, X_1),\ k = 0, \pm 1, \pm 2, ..., \tag{14}$$

where $\gamma^{(\mathrm{G})}(k)$ is simply the *Gini covariance* of X_{1+k} with respect to X_1 and in this context is called the *Gini covariance of lag k*. Of course, $\gamma^{(\mathrm{G})}(0) = \alpha(X) = \beta(X, X)$. For lags of magnitude $|k| \neq 0$, we have two separate measures of dependence, $\gamma^{(\mathrm{G})}(+|k|)$ and $\gamma^{(\mathrm{G})}(-|k|)$.

By way of contrast, under second order moment assumptions the usual autocovariance function $\gamma(k) = \mathrm{Cov}(X_{1+k}, X_1)$ is defined and satisfies $\gamma(k) = \gamma(-k)$. It is also of interest to compare $\{\gamma(\cdot)\}$ and $\{\gamma^{(\mathrm{G})}(\cdot)\}$ quantitatively, when both happen to be defined. This is best described by converting to correlations. The usual autocorrelation function is given by

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)},\ k = 0, \pm 1, \pm 2, ..., \tag{15}$$

while the *Gini autocorrelation function* is given by

$$\rho^{(\mathrm{G})}(k) = \frac{\gamma^{(\mathrm{G})}(k)}{\gamma^{(\mathrm{G})}(0)},\ k = 0, \pm 1, \pm 2, ... \tag{16}$$

Let us now add the common assumption that the stationary time series $\{X_t\}$ is *causal*, in which case, as discussed in the Appendix, we have a linear representation for X_t in terms of the innovation series $\{\varepsilon_s, s \leq t\}$:

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \tag{17}$$

where $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$, with $\psi(z) = \phi(z)^{-1}$ and $\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j$, and it is required that $\sum_{j=0}^{\infty} |\psi_j| < \infty$. (We find $\psi_0 = 1$, $\psi_1 = \phi_1$, $\psi_2 = \phi_2 + \phi_1^2$, etc.) Then (17) yields $\gamma(k) = \text{Var}(\varepsilon) \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}$ and $\rho(k) = \sum_{j=0}^{\infty} \psi_j \psi_{j+k} / \sum_{j=0}^{\infty} \psi_j^2$, $k \geq p$ (Brockwell and Davis, 1991).

This leads to an illuminating coherence between the Gini approach and the usual one. Let us define two new autocovariance functions, the *response-error (RE) autocovariance function*

$$\gamma^{(\text{RE})}(k) = \text{Cov}(X_{1+k}, \varepsilon_1), \ \ k = 0, 1, 2, \ldots, \tag{18}$$

defined under the usual second order moment assumptions, and the *Gini response-error (GRE) autocovariance function*

$$\gamma^{(\text{GRE})}(k) = \beta(X_{1+k}, \varepsilon_1), \ \ k = 0, 1, 2, \ldots, \tag{19}$$

defined under only first moment assumptions. For $k = 0$ these reduce to $\text{Var}(\varepsilon)$ and $\text{GMD}(\varepsilon)$, respectively, and for $k > 0$ they provide lag k measures of dependence between X_{1+k} and ε_1. (For $k < 0$ the quantities $\gamma^{(\text{RE})}(k)$ and $\gamma^{(\text{GRE})}(k)$ are simply 0 and thus not needed.) It is shown in the Appendix that when (17) holds we have

$$\gamma^{(\text{RE})}(k) = \psi_k \text{Var}(\varepsilon), \ \ k \geq 0, \tag{20}$$

$$\gamma^{(\text{GRE})}(k) = \psi_k \text{GMD}(\varepsilon), \ \ k \geq 0. \tag{21}$$

That is, when both defined, the two RE type autocovariance functions exhibit a strong coherence: they in common decay as ψ_k times the relevant measure of spread for ε, namely, respectively, the variance and the Gini mean difference of ε. We note in passing that a new interpretation for the quantities $\{\psi_k\}$ is provided by the notion of RE type autocovariance function, via (20) and (21), and that the Gini version is available under just first moment assumptions.

The above coherence between the usual autocovariance approach under second order moment assumptions and the new Gini autocovariance approach without those assumptions strongly motivates the Gini approach toward modeling of AR(p) in heavy tailed settings. This becomes reinforced in Section 3 where we introduce related new Gini-type Yule-Walker equations for the parameters $\phi_0, \ldots, \phi_p$.

In the special case of AR(1), an additional form of coherence holds. In the Appendix we show

Lemma A. *For AR*(1) *and* $k \geq 1$, X_{1+k} *has a linear regression on* X_1.

Moreover, under our stationarity assumption, X_{1+k} and X_1 have the same distribution F. Therefore, by Property (III) above, it follows that $\rho^{(\mathrm{G})}_{\mathrm{AR}(1)}(k) = \rho_{\mathrm{AR}(1)}(k)$, $k \geq 1$. That is, under second order moment assumptions, the AR(1) Gini autocorrelation function $\rho^{(\mathrm{G})}(k)$ agrees with the usual AR(1) autocorrelation function $\rho(k)$ for all $k \geq 1$ and provides additional information for $k < 0$. Moreover for AR(1), $\gamma(k) = \mathrm{Var}(\varepsilon)\phi_1^k/(1-\phi_1^2)$ (Brockwell and Davis, 1991), and it follows that

$$\rho^{(\mathrm{G})}_{\mathrm{AR}(1)}(k) = \rho_{\mathrm{AR}(1)}(k) = \phi_1^k,\ \ k \geq 1. \tag{22}$$

Thus for AR(1) there is an especially strong degree of coherence between the Gini and the usual autocovariance functions when both are defined.

2.2 Estimates based on a sample of independent observations

Although our main target is estimation of Gini mean difference, Gini covariance, and Gini autocovariance under the AR(p) model, it is helpful to begin with the case of a sample of *independent observations* and then introduce suitable modifications. Also, the case of independent observations will be relevant to our secondary aim of treating the MR model.

Sample Gini mean difference

Let $\{X_1, \ldots, X_n\}$ be a sample of independent observations having *univariate* distribution F, and denote the ordered values by $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$. The corresponding *sample distribution function* is defined by

$$\widehat{F}_n(x) = n^{-1}\sum_{i=1}^{n} I\{X_i \leq x\},\ \ -\infty < x < \infty,$$

where $I(\cdot)$ is the usual "indicator function", $I(A) = 1$ or 0 according as event A holds or not. The usual *sample Gini mean difference* for esimation of $\alpha(F)$ can be expressed in two different but algebraically equivalent forms,

$$\widehat{\alpha}_n = \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n} |X_i - X_j|, \tag{23}$$

which is a *U-statistic* in form, and

$$\widehat{\alpha}_n = \frac{2}{n(n-1)}\sum_{i=1}^{n}(2i-n-1)X_{i:n}, \tag{24}$$

an *L-statistic* in form. See Serfling (1980, 5.1.1 and 8.1.1) for discussion. An alternative, but quite natural approach, is to estimate $\alpha(F)$ by its sample analogue estimator $\alpha(\widehat{F}_n)$. Substitution of $\widehat{F}_n$ for F in (8) yields

$$\widehat{\alpha}^*_n = \alpha(\widehat{F}_n) = 2\int x\,(2\widehat{F}_n(x) - 1)\,d\widehat{F}_n(x) = \frac{2}{n^2}\sum_{i=1}^{n}(2i-n-1)X_{i:n}, \tag{25}$$

which is asymptotically equivalent to $\widehat{\alpha}_n$.

Immediately from (5) and (23) we see that $\widehat{\alpha}_n$ is *unbiased*: $E(\widehat{\alpha}_n) = \alpha(F)$. Also, because a U-statistic converges with probability 1 to its expected value (Serfling, 1980), $\widehat{\alpha}_n$ and $\widehat{\alpha}_n^*$ are *consistent* estimators of $\alpha(F)$. Their asymptotic normality is treated in Serfling (1980, 8.2.4) applying the L-statistic representation.

The asymptotic normality result depends, however, on the actual underlying distribution F, which is not assumed known, and, further, is based on $n \to \infty$. Therefore, instead of stating explicitly the details of the asymptotic normality, which are complicated, we recommend using the bootstrap method for inference purposes (see, for example Efron and Tibshirani, 1993, Davison and Hinkley, 1997, or Chernick, 2008).

Sample Gini covariance

Now let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent observations having a *bivariate* distribution F, and denote the concomitant of the ith ordered sample Y-value $Y_{i:n}$ by $X_{[i:n]}$, for $i = 1, \ldots, n$. Then (Schechtman and Yitzhaki, 1987, and Serfling and Xiao, 2007) an *unbiased*, *consistent*, and *asymptotically normal* estimator of $\beta(X, Y)$ is

$$\widehat{\beta}_n(X, Y) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} (X_{[i:n]} - X_{[j:n]}) = \frac{2}{n(n-1)} \sum_{i=1}^{n} (2i - n - 1) X_{[i:n]}, \tag{26}$$

similar to formulas for $\widehat{\alpha}_n$ but with the concomitants $X_{[i:n]}$ instead of the order statistics $X_{i:n}$. Likewise, $\widehat{\beta}_n(Y, X)$ is defined similarly in terms of the concomitants $Y_{[i:n]}$ of the ordered sample X-values.

2.3 Estimates in the stationary time series setting

Now let $\{X_1, X_2, \ldots, X_T\}$ be the observations over time $t = 1$ to $t = T$ of a stationary time series $\{X_t\}$ having common distribution F. Again, the sample distribution function $\widehat{F}_T$ defined above estimates the common distribution F, even though the sample values are now dependent, in the sense that $E\widehat{F}_T(x) = F(x)$ (unbiasedness) and, under a suitable rate of decay of the lag k covariances of $I\{X_1 \le x\}$ and $I\{X_{1+k} \le x\}$ as $k \to \infty$, $\widehat{F}_T(x)$ converges to $F(x)$ as $T \to \infty$. That is, we require that X_{1+k} become approximately independent of X_1 as $k \to \infty$ in the sense that $P(X_1 \le x, X_{1+k} \le x) \to P(X_1 \le x) \times P(X_{1+k} \le x)$.

It follows that $\gamma^{(\mathrm{G})}(0) = \alpha(F)$ is estimated by substitution of $\widehat{F}_T$ for F in (8), obtaining

$$\widehat{\gamma}_T^{(\mathrm{G})}(0) = 2 \int x \, (2\widehat{F}_T(x) - 1) \, d\widehat{F}_T(x) = \frac{2}{T^2} \sum_{i=1}^{T} (2i - T - 1) X_{i:T}, \tag{27}$$

parallel to $\widehat{\alpha}_n^*$ in (25). For estimation of $\gamma^{(\mathrm{G})}(k)$ for $k \ge +1$, let us use (10) to write

$$\gamma^{(\mathrm{G})}(k) = 2 \int \int x \, (2F(y) - 1) \, dF_{X_{1+k}, X_1}(x, y), \tag{28}$$

and let $\widehat{F}^{(k)}_{X_1}$ and $\widehat{F}^{(k)}_{X_{1+k},X_1}$ be sample versions of F_{X_1} and F_{X_{1+k},X_1} based on the $T-k$ lag k bivariate observations

$$\mathcal{S} = \{(X_{1+k}, X_1), (X_{2+k}, X_2), \ldots, (X_T, X_{T-k})\}.$$

Then substitution of these sample versions into (28) yields

$$\widehat{\gamma}^{(\mathrm{G})}_T(k) = \frac{2}{(T-k)^2}\sum_{t=1}^{T-k}(2t-(T-k)-1)X_{[t:T-k],1,2},\ k \geq +1, \tag{29}$$

where $X_{[t:T-k],1,2}$ is the *1st component* value that is concomitant to the tth ordered *2nd component* value, relative to the bivariate pairs of observations in $\mathcal{S}$. Similarly, for $k \leq -1$, and relative to the same set $\mathcal{S}$ of bivariate pairs, we derive

$$\widehat{\gamma}^{(\mathrm{G})}_T(k) = \frac{2}{(T-|k|)^2}\sum_{t=1}^{T-|k|}(2t-(T-|k|)-1)X_{[t:T-|k|],2,1},\ k \leq -1, \tag{30}$$

where $X_{[t:T-|k|],2,1}$ is the *2nd component* value that is concomitant to the tth ordered 1st component value.

3 A construction of linear systems using covariances

In general, given a linear system of p equations with coefficients given by certain population parameters $\phi_1, \ldots, \phi_p$ of special interest, a matrix inversion yields explicit formulas for those parameters. Consider the system of equations

$$a_i = \sum_{j=1}^{p} b_{ij}\phi_j,\ i = 1, \ldots, p, \tag{31}$$

or in matrix form

$$\boldsymbol{a} = \boldsymbol{B}\boldsymbol{\phi}, \tag{32}$$

with $\boldsymbol{a} = (a_1, \ldots, a_p)^T$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^T$, and $\boldsymbol{B} = (b_{ij})_{p\times p}$ (here $\boldsymbol{M}^T$ is the transpose of matrix $\boldsymbol{M}$). Relative to any other matrix $\boldsymbol{C} = (c_{ij})_{p\times p}$ that we might choose, for which $\boldsymbol{C}^T\boldsymbol{B}$ is invertible, there follows a corresponding solution for $\boldsymbol{\phi}$:

$$\boldsymbol{\phi} = (\boldsymbol{C}^T\boldsymbol{B})^{-1}\boldsymbol{C}^T\boldsymbol{a}. \tag{33}$$

The first challenge is to find choices of $\boldsymbol{a}$ and $\boldsymbol{B}$ which not only yield (32), but also are straightforward to evaluate and to estimate, and have intuitive appeal. The second challenge is to choose $\boldsymbol{C}$ strategically. When $\boldsymbol{B}$ is of rank p, the matrix $\boldsymbol{B}^T\boldsymbol{B}$ is invertible and thus simply taking $\boldsymbol{C} = \boldsymbol{B}$ in (33) yields the "usual" solution $\boldsymbol{\phi} = (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{a}$. But in general we are not confined to this choice of $\boldsymbol{C}$. We now introduce a simple device for obtaining suitable $\boldsymbol{a}$ and $\boldsymbol{B}$ via covariance methods, and then we apply the device to the AR and MR problems.

A simple device

We introduce a general covariance approach applicable when $\phi_1, \ldots, \phi_p$ are the coefficients in a linear regression structure,

$$\eta = \phi_0 + \sum_{j=1}^{p} \phi_j \xi_j + \varepsilon, \tag{34}$$

with ε independent of $\xi_1, \ldots, \xi_p$, as we are assuming in the AR and MR models. Immediately, for any function $Q(\xi_1, \ldots, \xi_p)$ of $\xi_1, \ldots, \xi_p$, we have

$$\mathrm{Cov}(\eta, Q(\xi_1, \ldots, \xi_p)) = \sum_{j=1}^{p} \phi_j \mathrm{Cov}(\xi_j, Q(\xi_1, \ldots, \xi_p)). \tag{35}$$

Then, using p different functions $Q_i(\xi_1, \ldots, \xi_p)$, $1 \leq i \leq p$, in (35), we obtain a linear system of equations of form (32), with

$$\begin{aligned} a_i &= \mathrm{Cov}(\eta, Q_i(\xi_1, \ldots, \xi_p), \ 1 \leq i \leq p, \\ b_{ij} &= \mathrm{Cov}(\xi_j, Q_i(\xi_1, \ldots, \xi_p)), \ 1 \leq i, j \leq p. \end{aligned}$$

In particular, in what follows we shall choose simply $Q_i(\xi_1, \ldots, \xi_p) = g(\xi_i)$, $1 \leq i \leq p$, for different choices of function g. In this case, for any such g, we obtain

$$a_i = \mathrm{Cov}(\eta, g(\xi_i)), \ 1 \leq i \leq p, \tag{36}$$

$$b_{ij} = \mathrm{Cov}(\xi_j, g(\xi_i)), \ 1 \leq i, j \leq p. \tag{37}$$

4 Gini autoregression approach

Let us take apply the device of Section 3 with $(\eta, \xi_1, \ldots, \xi_p) = (X_t, X_{t-1}, \ldots, X_{t-p})$, as per the AR($p$) model (1). We assume that $\{X_t\}$ is *causal* and *strictly stationary* with ε_t having *finite mean zero*, each t. As a reference point for comparison, we first examine the least squares approach and then introduce our Gini approach.

The standard least squares method

Use $g(\xi) = \xi$ and take $Q_i(X_{t-1}, \ldots, X_{t-p}) = X_{t-i}$, $1 \leq i \leq p$, yielding, under second order moment assumptions, $a_i = \mathrm{Cov}(X_t, X_{t-i}) = \gamma(i)$, $1 \leq i \leq p$, and $b_{ij} = \mathrm{Cov}(X_{t-j}, X_{t-i})$, $= \gamma(|i-j|)$, $1 \leq i, j \leq p$. In this case, (32) gives the usual "Yule-Walker equations" for $\phi_1, \ldots, \phi_p$ (e.g., Brockwell and Davis, 1991).

For $p = 1$ the least squares solution is simply $\phi_1 = \gamma(1)/\gamma(0)$, with estimator based on data

$$\widehat{\phi}_1 = \widehat{\gamma}_T(1)/\widehat{\gamma}_T(0) = \left(\frac{T}{T-1}\right)^2 \frac{\sum_{t=1}^{T-1}(X_{t+1} - \overline{X})(X_t - \overline{X})}{\sum_{t=1}^{T}(X_t - \overline{X})^2}. \tag{38}$$

A Gini covariance approach

We impose no moment assumptions higher than first order. Using $g(\xi) = 2(2F_X(\xi) - 1)$ and thus $Q_i(X_{t-1}, \ldots, X_{t-p}) = 2(2F_{X_{t-i}}(X_{t-i}) - 1)$, we obtain $a_i = \beta(X_t, X_{t-i}) = \gamma^{(\mathrm{G})}(i)$, $1 \leq i \leq p$, and $b_{ij} = \beta(X_{t-j}, X_{t-i}) = \gamma^{(\mathrm{G})}(i-j)$, $1 \leq i, j \leq p$. In this case, (32) gives what we call the "*Gini-Yule-Walker equations*" for $\phi_1, \ldots, \phi_p$.

For $p = 1$ the "Gini" solution is simply $\phi_1 = \gamma^{(\mathrm{G})}(1)/\gamma^{(\mathrm{G})}(0)$. From (27) and (29), its estimator based on data is

$$\widehat{\phi}_1^{(\mathrm{G})} = \widehat{\gamma}_T^{(\mathrm{G})}(1)/\widehat{\gamma}_T^{(\mathrm{G})}(0) = \left(\frac{T}{T-1}\right)^2 \frac{\sum_{t=1}^{T-1}(2t-(T-1)-1)X_{[t:T-1],1,2}}{\sum_{t=1}^{T}(2t-T-1)X_{t:T}}. \qquad (39)$$

A numerical illustration

As an illustration of how the Gini approach compares with the usual least squares approach, we generated the autoregression $X_t = 0.9X_{t-1} + \varepsilon_t$, $t = 1, \ldots, 50$, with two series of innovations $\{\varepsilon_t\}$, corresponding to a common distribution given by either a) a centered Pareto(1.5) or b) standard normal. The Pareto(θ) distribution considered here is that with cdf $F(x) = 1 - x^{-\theta}$, $x > 1$, and for the centered version we subtract the mean $\theta/(1-\theta)$, which is finite for $\theta > 1$. For our implementation with $\theta = 1.5$, the mean is finite but the variance is infinity. The resulting series $\{X_1, \ldots, X_{50}\}$ for the two cases are as follows:

```
Autoregression with centered Pareto innovations:

  -1.9483   -3.5950   -4.2585   -5.5537   -6.2821   -5.4694   -5.7306
  -7.0340   -8.0836   -3.4688   -5.0996   -0.9236   -2.5243   -3.5898
  -3.0987   -4.7179   -5.4553   -0.7672   -1.9233   -2.6715   -3.2025
  -4.5749   -4.0026   -5.4019   -6.7667   -2.0889   15.5084   12.3705
   9.1497    6.2545    6.2679    3.9921    1.9475   -0.0785    0.7932
  10.4619    8.0492    5.5961    3.3643    5.2102    2.9976    1.1755
   4.1073    2.5740    0.4976   -1.2926   -2.4167   -1.2632    0.2793
   2.6342
```

```
Autoregression with standard normal innovations:

  -0.0832539    0.4573340    1.8731149    0.8782000    0.8696156    1.2221884
   2.2844855    1.7871418    1.9879970    1.4837938    1.4051102    0.7361693
   0.1895652    0.0052900   -0.3288439   -0.4803858   -2.9699899   -2.5152607
  -4.7185150   -4.1156151   -4.0254369   -3.1347900   -2.6168699   -1.5996112
  -0.7694460   -1.8642228   -0.8086438    0.2295905    0.1968480   -0.0495835
  -0.9300178   -0.0535955   -1.4945776   -2.6791021   -2.5886898   -1.6836628
   0.5581806    1.0903191    1.0461148    0.7055263    2.5910885    0.9057135
  -0.4757917    0.2054224    0.7533263    0.0832891   -0.5738899    0.1223602
   0.1057207   -0.3673604
```

The goal is to estimate the lag 1 autocorrelation $\phi_1 = 0.9$. For the heavy tailed Pareto case, the standard least squares approach gives $\widehat{\phi}_1^{(\mathrm{LS})} = 0.78$, while the Gini approach gives $\widehat{\phi}_1^{(\mathrm{G})} = 0.85$. For the normal case, the standard least squares approach gives $\widehat{\phi}_1^{(\mathrm{LS})} = 0.84$, while the Gini approach gives $\widehat{\phi}_1^{(\mathrm{G})} = 0.88$. Thus in both instances the Gini approach happens to perform best. Of course, a detailed numerical study is desirable, but this is beyond the scope of the present paper.

5 Gini multiple regression approach

Several well-established methods for multiple regression impose moment assumptions of only first order or even none at all. The regression quantiles or least absolute deviation (LAD) method (e.g., see Koenker and Basset, 1978, Portnoy and Koenker, 1997, Koenker, 2005) and rank based methods (e.g., Jurečková and Sen, 1996) require no moment assumptions at all. In a variation on the LAD approach under an added first order moment assumption, the Gini mean difference (GMD) $E|\varepsilon_1 - \varepsilon_2|$ is minimized, thus eliminating the need of specifying a location parameter (Olkin and Yitzhaki, 1992, and Schechtman, Yitzhaki, and Pudalov, 2010). However, none of these alternatives to least squares yield closed form expressions or simply solve a linear system.

Here, using the device of Section 3 as for the AR problem, we derive in a direct way a suitable linear system for the parameters of the MR model (3), now based on the Gini covariances $\beta(Y, X_i)$ and $\beta(X_i, Y)$ for $1 \leq i \leq p$ and $\beta(X_i, X_j)$ for $1 \leq i, j \leq p$, that is, using the Gini covariance matrix of the random $(p+1)$-vector $(Y, X_1, \ldots, X_p)$. Thus the computational features of least squares are obtained while requiring only first order moment assumptions. Given a data set of independent and identically distributed $(p+1)$-variate observations $(Y_i, X_{1i}, \ldots, X_{pi})$ each satisfying (3), $1 \leq i \leq n$, the estimators $\widehat{\beta}(Y, X_i)$, $\widehat{\beta}(X_i, Y)$, and $\widehat{\beta}(X_i, X_j)$ constructed as per (26) are unbiased, consistent, and asymptotically normal. Even in the case of finite second order moments, these estimators are less sensitive to extreme observations in the data than the usual sample covariances.

We apply the device of Section 3 with $(\eta, \xi_1, \ldots, \xi_p) = (Z, Y_1, \ldots, Y_p)$, as per the MR model (3). We assume that ε has *finite mean* $E(\varepsilon) = 0$ and is *independent* of the variables $Y_1, \ldots, Y_p$. Again, we first examine the least squares approach and then introduce the Gini approach.

The least squares approach

With $g(\xi) = \xi$ and thus $Q_i(Y_1, \ldots, Y_p) = Y_i$, and under second order moment assumptions on ε, we obtain $a_i = \mathrm{Cov}(Z, Y_i)$, $1 \leq i \leq p$, and $b_{ij} = \mathrm{Cov}(Y_j, Y_i)$, $1 \leq i, j \leq p$. In this case, (32) gives the usual "normal equations".

For $p = 1$, and expressing the model as $Z = \phi_0 + \phi_1 Y + \varepsilon$, the solution is $\phi_1 = \mathrm{Cov}(Z, Y)/\mathrm{Cov}(Y, Y)$. The estimator based on data $\{(Z_1, Y_1), \ldots, (Z_n, Y_n)\}$, labeled so that

$Y_1 < \ldots < Y_n$, may be expressed as a weighted average of slopes,

$$\widehat{\phi}_1^{(\mathrm{LS})} = \sum_{i>j} w_{ij} s_{ij},$$

with slopes $s_{ij} = (Z_i - Z_j)/(Y_i - Y_j)$ and weights $w_{ij} = (Y_i - Y_j)^2 / \sum_{i>j}(Y_i - Y_j)^2$.

The Gini covariance approach

Taking $g(\xi) = 2(2F_X(\xi) - 1)$ and thus $Q_i(Y_1, \ldots, Y_p) = 2(2F_{Y_i}(Y_i) - 1)$, we obtain $a_i = \beta(Z, Y_i)$, $1 \le i \le p$, and $b_{ij} = \beta(Y_j, Y_i)$, $1 \le i, j \le p$. In this case, (32) gives immediately the same system of equations derived somewhat differently by Olkin and Yitzhaki (1992) and Schechtman, Yitzhaki, and Pudalov (2010), who also treat other types of "Gini" solutions and discuss related diagnostic and goodness-of-fit procedures.

For $p = 1$ the Gini solution is $\phi_1^{(\mathrm{G})} = \beta(Z, Y)/\beta(Y, Y)$, and the estimator based on data $\{(Z_1, Y_1), \ldots, (Z_1, Y_n)\}$ with $Y_1 < \ldots < Y_n$ may again be expressed as a weighted average of slopes,

$$\widehat{\phi}_1^{(\mathrm{G})} = \sum_{i>j} \widetilde{w}_{ij} s_{ij},$$

with s_{ij} as above but different weights $\widetilde{w}_{ij} = (Y_i - Y_j)/\sum_{i>j}(Y_i - Y_j)$. This facilitates comparison with the LS approach. However, more directly, we may use our estimators $\widehat{\beta}_n(Z, Y)$ and $\widehat{\beta}_n(Y, Y)$ to write

$$\widehat{\phi}_1^{(\mathrm{G})} = \frac{\widehat{\beta}_n(Z, Y)}{\widehat{\beta}_n(Y, Y)} = \frac{\sum_{i=1}^n (2i - n - 1) Z_{[i:n]}}{\sum_{i=1}^n (2i - n - 1) Y_{i:n}}, \tag{40}$$

with the concomitants $Z_{[i:n]}$ those of the ordered sample Y-values.

Appendix

Background for Equation (17). Following Brockwell and Davis (1991), an AR(p) model is a *causal* function of the innovations $\{\varepsilon_t\}$ if there exists a sequence of constants $\{\psi_t\}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \; t = 0, \pm 1, \ldots. \tag{41}$$

Application of this notion is based on the following general result.

> **Theorem**. *For any sequence of random variables $\{\varepsilon_t\}$ such that $\sup_t E|\varepsilon_t| < \infty$, and for any sequence of constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$, the series*
>
> $$\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \tag{42}$$

converges absolutely with probability 1. If also $\sup_t E|\varepsilon_t| < \infty$*, then the partial sums of the series in (34) converge in the mean to the same limit.*

This result is given under second order moment assumptions as Proposition 3.1.1 of Brockwell and Davis (1991). Their proof extends to the above more general setting with but minor modifications. The above assumption on $\{\varepsilon_t\}$ is met by our basic stationarity and first order moment assumptions, so that added the *causality* assumption does indeed yield (17), whose meaning is clarified by the above theorem. See Brockwell and Davis (1991) for detailed discussion.

Proof of Equations (20) and (21). Using (17), and applying independence of the innovations ε_t, we may write, for $k \geq 0$,

$$\begin{aligned}
\mathrm{Cov}(X_{1+k}, \varepsilon_1) &= \mathrm{Cov}\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{1+k-j},\, \varepsilon_1\right) \\
&= \sum_{j=0}^{\infty} \psi_j \mathrm{Cov}(\varepsilon_{1+k-j},\, \varepsilon_1) \\
&= \psi_k \mathrm{Var}(\varepsilon),
\end{aligned}$$

which gives (20). Similarly, again for $k \geq 0$,

$$\begin{aligned}
\beta(X_{1+k}, \varepsilon_1) &= \beta\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{1+k-j},\, \varepsilon_1\right) \\
&= 4\mathrm{Cov}\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{1+k-j},\, F_\varepsilon(\varepsilon_1)\right) \\
&= 4\sum_{j=0}^{\infty} \psi_j \mathrm{Cov}(\varepsilon_{1+k-j},\, F_\varepsilon(\varepsilon_1)) \\
&= \psi_k \mathrm{GMD}(\varepsilon),
\end{aligned}$$

giving (21).

Proof of Lemma A. For the model AR(1), we may write, for $k \geq 1$,

$$\begin{aligned}
X_{1+k} &= \phi_0 + \phi_1 X_k + \varepsilon_{1+k} \\
&= \phi_0 + \phi_1(\phi_0 + \phi_1 X_{k-1} + \varepsilon_k) + \varepsilon_{1+k} \\
&= \phi_0 + \phi_1\phi_0 + \phi_1^2 X_{k-1} + \phi_1\varepsilon_k + \varepsilon_{1+k} \\
&= \ldots \\
&= u(\phi_0, \phi_1) + \phi_1^k X_1 + v(\phi_0, \phi_1, \varepsilon_2, \ldots, \varepsilon_k, \varepsilon_{1+k}),
\end{aligned}$$

for some functions $u(\cdot)$ and $v(\cdot)$. Since $v(\phi_0, \phi_1, \varepsilon_2, \ldots, \varepsilon_k, \varepsilon_{1+k})$ is a linear combination of the variables $\varepsilon_2, \ldots, \varepsilon_k, \varepsilon_{1+k}$, which are independent of X_1 and have mean 0, it follows that

$$E(X_{1+k} \mid X_1) = u(\phi_0, \phi_1) + \phi_1^k X_1.$$

That is, X_{1+k} has a linear regression on X_1. □

Acknowledgements

The author greatly appreciates Jed Frees kindly providing a pre-publication draft of his recent book [12]. Also, support under a Society of Actuaries Grant, National Science Foundation Grants DMS-0103698 and DMS-0805786, and NSA Grant H98230-08-1-0106, is gratefully acknowledged.

References

[1] Bloomfield, P. and Steiger, W. S. (1983). *Least Absolute Deviations: Theory, Applications and Algorithms*. Birkhaüser, Boston.

[2] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd edition. Springer.

[3] Chernick, M. R. (2008) *Bootstrap Methods: A Guide for Practitioners and Researchers*, 2nd edition. Wiley.

[4] Daniels, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika* **33** 129–135.

[5] David, H. A. and Nagaraja, H. N. (2003). *Order Statistics*, 3rd Edition. Wiley, New York.

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

[7] Davis, R. A. and Resnick, S. I. (1985b). More limit theory for the sample correlation function of moving averages. *Stochastic Processes and Their Applications* **30** 257–259.

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

[9] Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application.* Cambridge University Press.

[10] Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman & Hall.

[11] Feigin, P. D. and Resnick, S. I. (1999). Pitfalls of fitting autoregressive models for heavy-tailed time series. *Extremes* **1** 391-422.

[12] Frees, E. W. (2010). *Regression Modeling with Actuarial and Financial Applications*. Springer.

[13] Gini, C. (1912). Variabilità e mutabilità, contributio allo studio delle distribuzioni e delle relazione statistiche. *Studi Economico-Giuridici della Reale Università di Cagliari* **3** 3–159.

[14] Jurečková, J. and Sen, P. K. (1996). em Robust Statistical Procedures: Asymptotics and Interrelations. John Wiley, New York.

[15] Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, New York.

[16] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Journal of the American Statistical Association* **82** 851–857.

[17] Klugman, Stuart A., Harry H. Panjer and Gordon E. Willmot (2004). *Loss Models: From Data to Decisions*, 2nd edition. Wiley.

[18] Ling, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society*, *Series B* **67** 381–393.

[19] Olkin, I. and Yitzhaki, S. (1992). Gini regression analysis. *International Statistical Review* **60** 185–196.

[20] Portnoy, S. and Koenker, R. (1997). The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error versus Absolute-Error Estimators, with discussion. *Statistical Science* **12** 279–300.

[21] Resnick, S. I. (1997). Heavy tail modeling and teletraffic data (with discussion). *Annals of Statistics* **25** 1805–1869.

[22] Rosenberg, M. A. and Farrell, P. M. (2008). Predictive modeling of costs for a chronic disease with acute high cost episodes. *North American Actuarial Journal* **12** 1–18.

[23] Rosenberg, M. A., Frees, E. W., Sun, J. , Johnson, P., and Robinson, J. M. (2007). Predictive modeling with longitudinal data: A case study of Wisconsin nursing homes. *North American Actuarial Journal* **11** 54–69.

[24] Schechtman, E. and Yitzhaki, S. (1987). A measure of association based on Gini's mean difference. *Communications in Statistics – Theory and Methods* **16** 207–231.

[25] Schechtman, E., Yitzhaki, S., and Artsev, Y. (2008). Who does not respond in the household expenditure survey: an exercise in extended Gini regressions. *Journal of Business & Economic Statistics* **26** 329–344.

[26] Schechtman, E., Yitzhaki, S., and Pudalov, T. (2010). Gini's multiple regressions: two approaches and their interaction. Preprint. Available at SRRN: http://ssrn.com/abstract=1626803.

[27] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.

[28] Serfling, R. and Xiao, P. (2007). A contribution to multivariate L-moments: L-comoment matrices. *Journal of Multivariate Analysis* **98** 1765–1781.

[29] Yitzhaki, S. and Olkin, I. (1991). Concentration indices and concentration curves. In *Stochastic Orders and Decision under Risk* (K. Mosler and M. Scarsini, eds.), pp. 380–392. IMS Lecture Notes – Monograph Series, Volume 19, Hayward, California.