Statistical Inference for Lorenz Curves with Censored Data

by

Frank A Cowell
STICERD, London School of Economics and Political Science

and

Maria-Pia Victoria-Feser
Université de Genève

Discussion Paper
No.DARP/35
June 1998

The Toyota Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020-7955 6678

The first author is partially supported by ESRC grants R000235725 and R5550XY8. The second author is partially supported by the 'Fond National Suisse pour la Recherche Scientifique. We are grateful to Elvezio Ronchetti and Christian Schluter for helpful comments, and to Kaspar Richter and Hung Wai Wong for research assistance.

Abstract

Lorenz curves and associated tools for ranking income distributions are commonly estimated on the assumption that full, unbiased samples are available. However, it is common to find income and wealth distributions that are routinely censored or trimmed. We derive the sampling distribution for a key family of statistics in the case where data have been modified in this fashion.

Keywords: Lorenz curve; sampling errors.

JEL Nos.: C13, D63.

Contact address: Contact address: Professor Frank Cowell, STICERD, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK. Email: f.cowell@lse.ac.uk

1 Introduction

The use of Lorenz comparisons in distributional analysis is viewed by many as fundamental. In order to be able to implement the principles of such comparisons in practice it is necessary to have appropriate statistical tools. Building on a classic paper by Beach and Davidson it has become standard practice to apply statistical tests to distributional comparisons and to construct confidence bands for Lorenz curves and associated tools (Beach and Davidson 1983) (Beach and Kaliski 1986) (Beach and Richmond 1985).[1] However, although the results in this literature are "distribution-free", in that they impose few restrictions on the assumed underlying distribution in the population, they do rest upon some fairly demanding assumptions about the sample - which may be systematically violated in many practical applications to income- and wealth-distributions. In this paper we establish results for a more general approach which takes into account some of these difficulties.

It is now commonly recognised that income and wealth data may have been censored or trimmed for reasons of confidentiality or convenience (Fichtenbaum and Shahidi 1988). This treatment of the data means that the conventional analysis - reviewed in Section 3 below - is not applicable. Section 4 derives analogous results for censored data and shows their relationship to the standard

[1] For other applications of classical hypothesis testing to ranking criteria see, for example, Beach et al. (1994), **?** (?, ?, ?, ?), Bishop et al. (1991), ?), Davidson and Duclos (1997), Stein et al. (1987).

literature.

Of course it should be recognised that deriving asymptotic distributions and moments is just one possible way of providing the statistical tools for testing distributional comparisons. So, before presenting our results, a word should be said about other ways of making inference from sample data. An alternative approach would be to use the bootstrap (Efron 1979, Efron and Tibshirani 1993) which is particularly suitable in the case of small samples. However, the bootstrap does not work in every case, especially when the statistic to be bootstrapped is bounded, as Schenker (1985) and Andrews (1997) have shown. Applications to Lorenz curves might fail in these special cases, but to prove this is beyond the scope of the present paper.

2 Terminology and Notation

Let $\mathfrak{F}$ be the set of all continuous probability distributions and $\mathfrak{F}^*$ the subset of $\mathfrak{F}$ which is twice differentiable with finite mean and variance. Let income be a continuous random variable X with probability distribution $F \in \mathfrak{F}$ and support on the interval $[\underline{x}, \bar{x}] \subseteq \Re$, where $\Re$ is the real line. Write the mean of the distribution as the functional $\mu(F)$.

We require three fundamental functionals from $\mathfrak{F} \times [0,1]$ to $\Re$. Let $q \in [0,1]$

denote an arbitrary population proportion: the *quantile functional* is defined by:

$$Q(F;q) := \inf\{x|F(x) \geq q\} =: x_q \tag{1}$$

(Gastwirth 1971), and the *cumulative income functional* is defined by:

$$C(F;q) := \int_{\underline{x}}^{x_q} x dF(x) =: c_q \tag{2}$$

(Cowell and Victoria-Feser 1996); analogously define:

$$S(F;q) := \int_{\underline{x}}^{x_q} x^2 dF(x) =: s_q. \tag{3}$$

The functional (2) is used to define the following standard concepts. For a given $F \in \mathfrak{F}$, the graph $\{q, C(F,q)\}$ describes the *generalised Lorenz curve* (GLC). The scale normalisation of the GLC by the mean gives the (relative) Lorenz functional:

$$L(F;q) := \frac{C(F;q)}{\mu(F)} \tag{4}$$

and the graph $\{q, L(F;q)\}$ gives the *relative Lorenz curve.*[2] An alternative normalisation of the GLC yields the absolute counterpart to (4)

[2]Beach and Davidson use a different, related concept to underpin these expressions, the first moment function $\Phi : \mathfrak{X} \mapsto [0,1]$ given by $\Phi(x) = L(F;x) = \frac{1}{\mu(F)} \int^x y dF(y)$ (Kendall and Stuart 1977).

$$A(F;q) := C(F;q) - q\mu(F) \tag{5}$$

and the graph $\{q, A(F;q)\}$ is the *absolute Lorenz Curve* (Moyes 1987).

In estimating the quantiles and income cumulations one uses $F^{(n)}$, a sample of size n drawn from the distribution F; this is a distribution consisting of n point-masses $\frac{1}{n}$, one at each observation in the sample. Denote the order statistics of the sample by $\left\{x_{[i]} : i = 1, ..., n\right\}$. How these are to be implemented appropriately depends upon the nature of the sample as we discuss in Sections 3 and 4.

3 Empirical Implementation: Uncensored Samples

Choose a finite collection of population proportions $\Theta \subset [0,1]$; then for each $q \in \Theta$ on can compute the sample quantiles and cumulants. Let $\text{int}(z)$ be the largest integer less than or equal to z, and let

$$\iota(n,q) := \text{int}[(n-1)q + 1] \tag{6}$$

denote the order of the observation corresponding to the quantile q. Then we have

$$\hat{x}_q := Q\left(F^{(n)}; q\right) = x_{[\iota(n,q)]} \tag{7}$$

$$\hat{c}_q := C\left(F^{(n)}; q\right) = \frac{1}{n}\sum_{i=1}^{\iota(n,q)} x_{[i]} \tag{8}$$

using (2). The set of pairs $\{(q, \hat{c}_q) : q \in \Theta\}$ gives points on the empirical generalised Lorenz curve, and $\hat{c}_1$ is the sample mean $\mu\left(F^{(n)}\right)$; the relative and absolute Lorenz curves are found by normalisation as in (4) and (5).

Following Beach and Davidson (1983) assume that the underlying distribution $F \in \mathfrak{F}^*$; then the slightly modified Theorem 1 in Beach and Davidson (1983) is:

Theorem 1 *The set of sample income cumulations* $\{\hat{c}_q : q \in \Theta\}$ *is asymptotically multivariate normal. For any* $q, q' \in \Theta$ *such that* $q \leq q'$ *the asymptotic covariance of* $\sqrt{n}\hat{c}_q$ *and* $\sqrt{n}\hat{c}_{q'}$ *is*

$$\omega_{qq'} := s_q + [qx_q - c_q]\left[x_{q'} - q'x_{q'} + c_{q'}\right] - x_q c_q \tag{9}$$

The proof provided in Beach and Davidson (1983) contains some slips which will obscure the relationship between the standard Beach-Davidson result and generalisations of it. For a corrected version see Appendix A.1.

Given Theorem 1 the sampling distribution of empirical representations of the GLC, RLC and the ALC are immediately obtained by standard methods. For the implementation of each concept one can readily set up appropriate confidence intervals and tests. However, all of them rest, of course, on the strong assumption that the data with which one is working consist of an unbiased sample from the population.

4 Censored Data

One major departure from the unbiased-sample case is that where the sample has been *censored*. Censored data are commonly encountered in practical applications to income and wealth distributions, for several reasons.

As noted in the introduction the data may have been censored by the data providers. This usually affects extreme values - the problem of "top-coding" and "bottom-coding". Some high-income observations may have been removed from the sample because of concern for confidentiality - in the sparse upper tail it might not be too difficult from a skilful data-detective to identify the individuals or households corresponding to individual observations. Some low-income observations may have been removed or modified for reasons of convenience: it has been known for zero or negative income values to be filtered from a sample before releasing it; the tax authorities may not find it practical or worthwhile to collect data from those of modest means. Sometimes extreme values have been modified because it was inconvenient to store large numbers in the data set.

Furthermore, because income cumulations are inherently non-robust statistics (Cowell and Victoria-Feser 1996) the prudent user of income or wealth statistics may selectively "trim" the sample data in order to obtain Lorenz-type comparisons that are not driven by extreme values (**?**).

We treat the problem of trimming or censoring by supposing that a given proportion of the observations has been removed from each end of the distribution.

Specifically we assume that a proportion $\underline{\alpha}$ has been removed from the bottom and $\bar{\alpha}$ from the top of the distribution, and define $\alpha := \underline{\alpha} + \bar{\alpha}$. Then

$$\mu_\alpha(F) = \int_{\underline{x}}^{\overline{x}} a(u) u dF(u) \tag{10}$$

with

$$a(u) := \begin{cases} \frac{1}{1-\alpha} & Q(F;\underline{\alpha}) < u \leq Q(F;\bar{\alpha}) \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

such that

$$\int_{\underline{x}}^{\overline{x}} a(u) dF(u) = 1$$

Assume that the set of population proportions satisfies $\Theta \subset [\underline{\alpha}, \bar{\alpha}]$. The quantile associated with any $q \in \Theta$ is $x_q = Q(F;q)$, so that the α-trimmed income cumulations are

$$C_\alpha(F;q) = \int_{\underline{x}}^{x_q} a(u) u dF(u) = \frac{1}{1-\alpha} \int_{x_{\underline{\alpha}}}^{x_q} u dF(u) =: c_{\alpha,q} \tag{12}$$

Alternatively, we have

$$\begin{aligned} C_\alpha(F;q) &= \frac{1}{1-\alpha}\left[\int_{\underline{x}}^{x_q} u dF(u) - \int_{\underline{x}}^{x_{\underline{\alpha}}} u dF(u)\right] && (13) \\ &= \frac{C(F;q) - C(F;\underline{\alpha})}{1-\alpha}. && (14) \end{aligned}$$

Also define

$$S_\alpha(F;q) := \int_{\underline{x}}^{x_q} a(u)u^2 dF(u) = \frac{1}{1-\alpha}\int_{x_{\underline{\alpha}}}^{x_q} u^2 dF(u) =: s_{\alpha,q}. \qquad (15)$$

The sample version is obtained by considering the sample quantiles (7) with n being the untrimmed sample size, and the first moment sample cumulants

$$\hat{c}_{\alpha,q} := C_\alpha\left(F^{(n)};q\right) = \frac{1}{n[1-\alpha]}\sum_{i=\iota(n,\underline{\alpha})+1}^{\iota(n,q)} x_{[i]} = \frac{\hat{c}_q - \hat{c}_{\underline{\alpha}}}{1-\alpha} \qquad (16)$$

$$= \frac{\hat{c}_q - \hat{c}_{\underline{\underline{\alpha}}}}{1-\alpha}$$

- see (6). Note that $x_{[\iota(n,\underline{\alpha})+1]}$ is the first observation in the trimmed sample. It is easy to show that

$$E[\hat{c}_{\alpha,q}] = \frac{c_\alpha - c_q}{1-\alpha} \qquad (17)$$

For the covariance, we need to find $[1-\alpha]^{-2}$ times

$$\operatorname{cov}\left(\sqrt{n}\left[\hat{c}_q - \hat{c}_{\underline{\alpha}}\right], \sqrt{n}\left[\hat{c}_{q'} - \hat{c}_{\underline{\alpha}}\right]\right) \qquad (18)$$

Expression (18) is equivalent to

$$n\ \left[\ \operatorname{cov}\left(\hat{c}_q, \hat{c}_{q'}\right) - \operatorname{cov}\left(\hat{c}_q, \hat{c}_{\underline{\alpha}}\right) - \operatorname{cov}\left(\hat{c}_{q'}, \hat{c}_{\underline{\alpha}}\right) + \operatorname{cov}\left(\hat{c}_{\underline{\alpha}}, \hat{c}_{\underline{\alpha}}\right)\right]$$

$$= \omega_{qq'} - s_{\underline{\alpha}} + x_{\underline{\alpha}} c_{\underline{\alpha}}$$

$$+\left[\underline{\alpha}x_{\underline{\alpha}}-c_{\underline{\alpha}}\right]\left[x_{\underline{\alpha}}-x_q-x_{q'}-\underline{\alpha}x_{\underline{\alpha}}+qx_q+q'x_{q'}+c_{\underline{\alpha}}-c_q-c_{q'}\right] \quad (19)$$

where $\omega_{qq'}$ is given by (9).

However the expression (19) is unsatisfactory, because an estimate of it would have to be based on the sample analogues of c_q, $c_{q'}$ and s_q which would need to be calculated using the whole sample (before censoring). We therefore need an expression that depends only on $c_{\alpha,q}$ and $s_{\alpha,q}$. We can do that by calculating directly (up to the multiplicative constant of $[1-\alpha]^{-2}$)

$$\operatorname{cov}\left(\sqrt{n}\hat{c}_{\alpha,q},\sqrt{n}\hat{c}_{\alpha,q'}\right). \quad (20)$$

As with the untrimmed case (see 33 in the Appendix), this covariance (20) can be expressed as

$$\int_{x_{\underline{\alpha}}}^{x_q}[1-F(y)]\int_{x_{\underline{\alpha}}}^{y}F(u)du\,dy+\int_{x_{\underline{\alpha}}}^{x_q}F(y)\int_{y}^{x_{q'}}[1-F(u)]\,du\,dy. \quad (21)$$

and (21) may then be rearranged to give (see 39 to 40 in the Appendix):

$$\int_{x_{\underline{\alpha}}}^{x_q}F(y)dy\int_{x_{\underline{\alpha}}}^{x_{q'}}[1-F(u)]\,du-\underline{\alpha}x_{\underline{\alpha}}\left[x_q-x_{\underline{\alpha}}\right]-\int_{x_{\underline{\alpha}}}^{x_q}\int_{x_{\underline{\alpha}}}^{y}udF(u)dy+x_{\underline{\alpha}}\int_{x_{\underline{\alpha}}}^{x_q}F(y)dy \quad (22)$$

Taking each integral in (22) separately, we have:

$$\int_{x_{\underline{\alpha}}}^{x_q} F(y)dy = [yF(y)]_{x_{\underline{\alpha}}}^{x_q} - \int_{x_{\underline{\alpha}}}^{x_q} ydF(y) = qx_q - \underline{\alpha}x_{\underline{\alpha}} - [1-\alpha]\, c_{\alpha,q} \tag{23}$$

$$\int_{x_{\underline{\alpha}}}^{x_{q'}} (1-F(u))du = [1-q']\, x_{q'} - [1-\underline{\alpha}]\, x_{\underline{\alpha}} + [1-\alpha]\, c_{\alpha,q'} \tag{24}$$

$$\begin{aligned} \int_{x_{\underline{\alpha}}}^{x_q}\int_{x_{\underline{\alpha}}}^{y} udF(u)dy &= \left[y\int_{x_{\underline{\alpha}}}^{y} udF(u)\right]_{x_{\underline{\alpha}}}^{x_q} - \int_{x_{\underline{\alpha}}}^{x_q} y^2 dF(y) \\ &= x_q \int_{x_{\underline{\alpha}}}^{x_q} udF(u) - \int_{x_{\underline{\alpha}}}^{x_q} y^2 dF(y) \\ &= x_q\,[1-\alpha]\, c_{\alpha,q} - [1-\alpha]\, s_{\alpha,q} \end{aligned} \tag{25}$$

$$x_{\underline{\alpha}} \int_{x_{\underline{\alpha}}}^{x_q} F(y)dy = x_{\underline{\alpha}}\,[qx_q - \underline{\alpha}x_{\underline{\alpha}} - [1-\alpha]\, c_{\alpha,q}] \tag{26}$$

So that, by substituting (23)-(26) in (22), we get

$$\begin{aligned} & [qx_q - \underline{\alpha}x_{\underline{\alpha}} - [1-\alpha]\, c_{\alpha,q}]\, [[1-q']\, x_{q'} - [1-\underline{\alpha}]\, x_{\underline{\alpha}} + [1-\alpha]\, c_{\alpha,q'}] - \\ & [x_q\,[1-\alpha]\, c_{\alpha,q} - [1-\alpha]\, s_{\alpha,q}] + x_{\underline{\alpha}}\,[qx_q - \underline{\alpha}x_{\underline{\alpha}} - [1-\alpha]\, c_{\alpha,q}] \\ = & \ (1-\alpha)^2 \varpi_{qq'} \end{aligned} \tag{27}$$

It can be shown that (27) is equal to (19). However, (27) is based on trimmed moments and the sample analogue will be also based on trimmed samples. This is both a more natural way of calculating the variances and an attractive empirical construct in that it is independent of detailed information censored from the distribution. So we may state:[3]

Theorem 2 *Given a sample of size* n *and lower and upper trimming proportions* $\underline{\alpha}, \bar{\alpha} \in [0,1]$ *the set of sample income cumulations* $\{\hat{c}_{\alpha,q} : q \in \Theta\}$ *is asymptotically multivariate normal. For any* $q, q' \in \Theta$ *such that* $q \leq q'$ *the asymptotic covariance of* $\sqrt{n}\hat{c}_{\alpha,q}$ *and* $\sqrt{n}\hat{c}_{\alpha,q'}$ *is given by* $\varpi_{qq'}$ *defined in (27).*

Theorem 2 can immediately be applied to the construction of practical tools for distributional comparisons. For example the 95% confidence interval for ordinates of the GLC will be given by

$$\hat{c}_{\alpha,q} \pm 1.96 n^{-1/2} \hat{\varpi}_{qq} \tag{28}$$

where $\hat{\varpi}_{qq}$ are the sample analogues of ϖ_{qq}.

For the relative Lorenz curves, we need to specify the following quantities. First note that

$$\hat{\mu}_\alpha = \hat{c}_{\alpha,1} = \hat{c}_{\alpha,1-\overline{\alpha}} \tag{29}$$

Therefore, the asymptotic variance of $\hat{\mu}_\alpha$ and the covariances between $\hat{\mu}_\alpha$ and $\hat{c}_{\alpha,q}$ are found using (19) or (27). Let the latter be denoted respectively by ϖ_{11} and ϖ_{q1}. Using the standard result on limiting distributions of differentiable functions

[3] Asymptotic normality follows from the fact that the required statistics are linear functions of order statistics - see Moore (1968), Shorack (1972), Stigler (1969, 1974).

of random variables (Rao 1973), the asymptotic covariances of the RLC ordinates are then given by

$$v_{qq'}^{\mathrm{RLC}} = \frac{1}{\mu_\alpha^2}\left[\varpi_{qq'} + c_{\alpha,q}c_{\alpha,q'}\varpi_{11} - c_{\alpha,q}\varpi_{q'1} - c_{\alpha,q'}\varpi_{q1}\right] .$$

Therefore, a 95% confidence interval for the ordinates of the RLC is given by

$$L(F^{(n)};q) \pm 1.96 n^{-1/2}\hat{v}_{qq}^{\mathrm{RLC}}$$

where $\hat{v}_{qq}^{RLC}$ is the sample analogue of v_{qq}^{RLC}.

For absolute Lorenz curves, following the same argument, we find that the asymptotic covariances of the ALC ordinates are then given by

$$v_{qq'}^{\mathrm{ALC}} = \varpi_{qq'} + qq'\varpi_{11} - q\varpi_{q'1} - q'\varpi_{q1} .$$

Therefore, a 95% confidence interval for the ordinates of the ALC is given by

$$A(F^{(n)};q) \pm 1.96 n^{-1/2}\hat{v}_{qq}^{\mathrm{ALC}} .$$

5 Concluding Remarks

Statistical inference with trimmed or censored Lorenz curves should be part of the standard repertoire of applied economists and statisticians working with income and wealth data. Our approach makes relatively few demands on the data: all that one needs to know is the amount by which the sample has been trimmed in each tail before the data are analysed. Furthermore the main result requires only modest (but important) extensions to the standard procedures in the literature. The analysis is straightforward to interpret and to translate into practical algorithms.

References

Andrews, D. W. K. (1997). A simple counterpart to the bootstrap. Discussion Paper 1157, Cowles Foundation for Research in Economics at Yale University, New Haven, Connecticut 06520-8281.

Beach, C. M., K. Chow, J. Formby, and G. Slotsve (1994). Statistical inference for decile means. *Economics Letters 45*, 161–167.

Beach, C. M. and R. Davidson (1983). Distribution-free statistical inference with Lorenz curves and income shares. *Review of Economic Studies 50*, 723–725.

Beach, C. M. and S. F. Kaliski (1986). Lorenz Curve inference with sample weights: an application to the distribution of unemployment experience. *Applied Statistics 35*(1), 38–45.

Beach, C. M. and J. Richmond (1985). Joint confidence intervals for income shares and Lorenz curves. *International Economic Review 26*(6), 439–450.

Bishop, J. A., J. P. Formby, and W. P. Smith (1991). International comparisons of income inequality: Tests for Lorenz dominance across nine countries. *Economica 58*, 461–477.

Cowell, F. A. and M.-P. Victoria-Feser (1996). Welfare judgements in the presence of contaminated data. Distributional Analysis Discussion Paper 13, STICERD, London School of Economics, London WC2A 2AE.

Davidson, R. and J.-Y. Duclos (1997). Statistical inference for the measurement of the incidence of taxes and transfers. *Econometrica 65*. forthcoming.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics 7*, 1–26.

Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap*. London: Chapman and Hall.

Fichtenbaum, R. and H. Shahidi (1988). Truncation bias and the measurement of income inequality. *Journal of Business and Economic Statistics 6*, 335–337.

Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica 39*, 1037–1039.

Kendall, M. and A. Stuart (1977). *The Advanced Theory of Statistics*. London: Griffin.

Moore, D. S. (1968). An elementary proof of asymptotic normality of linear functions of order statistics. *Annals of Mathematical Statistics 39*, 263–265.

Moyes, P. (1987). A new concept of Lorenz domination. *Economics Letters 23*, 203–207.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. New York: Wiley.

Schenker, N. (1985). Qualms about bootstrap confidence intervals. *Journal of the American Statistical Association 80*, 360–361.

Shorack, G. R. (1972). Functions of order statistics. *Annals of Mathematical Statistics 43*, 412–427.

Stein, W. E., R. C. Pfaffenberger, and D. W. French (1987). Sampling error in first-order stochastic dominance. *Journal of Financial Research 10*, 259–269.

Stigler, S. M. (1969). Linear functions of order statistics. *Annals of Mathematical Statistics 40*, 770–788.

Stigler, S. M. (1974). Linear functions of order statistics with smooth weight functions. *Annals of Statistics 2*, 676–693. "Correction", 7, 466.

A Proofs of Theorems

A.1 Uncensored Data (Beach-Davidson)

First note that the vector of sample statistics used in Theorem 1 of Beach and Davidson (1983) is, in our notation, $[\hat{c}_q]_{q\in\Theta}$. where $\hat{c}_q$ is given by (8) and $\Theta \subset [0,1]$.

Applying the results of Rao (1973), pp 387-388 and using (8) we have [4]

$$n \operatorname{cov}(\hat{c}_q, \hat{c}_{q'}) = \omega_{qq'} \tag{30}$$

where

$$\begin{aligned}
\omega_{qq'} \quad : \quad &= \int_0^q (1-p) \int_0^p t \frac{dQ(F;t)}{dt} \frac{dQ(F;p)}{dp} dt dp + && (31)\\
&\int_0^q p \int_p^{q'} (1-t) \frac{dQ(F;t)}{dt} \frac{dQ(F;p)}{dp} dt dp && (32)\\
&= \int_{\underline{x}}^{x_q} (1-F(y)) \int_{\underline{x}}^{y} F(u) du dy + \int_{\underline{x}}^{x_q} F(y) \int_y^{x_{q'}} (1-F(u)) du dy, && (33)
\end{aligned}$$

$p := F(y)$ and $t := F(u)$. Expression (33) yields

$$\int_{\underline{x}}^{x_q} \int_{\underline{x}}^{y} F(u) du dy - \int_{\underline{x}}^{x_q} F(y) \int_{\underline{x}}^{y} F(u) du dy + \\
\int_{\underline{x}}^{x_q} F(y) \int_y^{x_{q'}} du dy - \int_{\underline{x}}^{x_q} F(y) \int_y^{x_{q'}} F(u) du dy$$

[4]Expression (33) is equivalent to Beach and Davidson (1983), top of page 728 for the case where $\underline{x} = 0$.

$$= \int_{\underline{x}}^{x_q} \int_{\underline{x}}^{y} F(u)dudy + \int_{\underline{x}}^{x_q} F(y) \int_{y}^{x_{q'}} dudy - \int_{\underline{x}}^{x_q} F(y) \int_{\underline{x}}^{x_{q'}} F(u)dudy$$

$$\begin{aligned} = \; & \int_{\underline{x}}^{x_q} \int_{\underline{x}}^{y} F(u)dudy + \int_{\underline{x}}^{x_q} F(y) \int_{y}^{x_{q'}} dudy + \\ & \int_{\underline{x}}^{x_q} F(y)dy \int_{\underline{x}}^{x_{q'}} (1 - F(u))du - \int_{\underline{x}}^{x_q} F(y)dy \int_{\underline{x}}^{x_{q'}} du \end{aligned}$$

$$= \int_{\underline{x}}^{x_q} F(y)dy \int_{\underline{x}}^{x_{q'}} (1 - F(u))du - \int_{\underline{x}}^{x_q} F(y) \int_{\underline{x}}^{y} dudy + \int_{\underline{x}}^{x_q} \int_{\underline{x}}^{y} F(u)dudy$$

$$\begin{aligned} = \; & \int_{\underline{x}}^{x_q} F(y)dy \int_{\underline{x}}^{x_{q'}} (1 - F(u))du - \int_{\underline{x}}^{x_q} F(y) \left[y - \underline{x}\right] dy + \\ & \int_{\underline{x}}^{x_q} \left\{ \left[uF(u)\right]_{\underline{x}}^{y} - \int_{\underline{x}}^{y} udF(u) \right\} dy \end{aligned}$$

$$= \int_{\underline{x}}^{x_q} F(y)dy \left[\int_{\underline{x}}^{x_{q'}} (1 - F(u))du + \underline{x} \right] - \int_{\underline{x}}^{x_q} \int_{\underline{x}}^{y} udF(u)dy \qquad (34)$$

Evaluating the three main components of (34) we have:

$$\int_{\underline{x}}^{x_q} F(y)dy = \left[yF(y)\right]_{\underline{x}}^{x_q} - \int_{\underline{x}}^{x_q} ydF(y) = qx_q - c_q \qquad (35)$$

$$\int_{\underline{x}}^{x_{q'}} (1 - F(u))du + \underline{x} = x_{q'}(1 - q') + c_{q'} \qquad (36)$$

$$\int_{\underline{x}}^{x_q} \int_{\underline{x}}^{y} udF(u)dy = \left[y \int_{\underline{x}}^{y} udF(u) \right]_{\underline{x}}^{x_q} - \int_{\underline{x}}^{x_q} y^2 dF(y) = x_q c_q - s_q \qquad (37)$$

Substituting from (35)-(37) into (34) we find

$$\omega_{qq'} := \left[qx_q - c_q\right] \left[x_{q'}(1 - q') + c_{q'}\right] - \left[x_q c_q - s_q\right]$$

which is the result in (9).

A.2 Censored Data

To establish (21) in the analysis leading to Theorem 2 note that (20) may be written

$$\int_{\underline{\alpha}}^{q} [1-p] \int_{\underline{\alpha}}^{p} t \frac{dQ(F;t)}{dt} \frac{dQ(F;p)}{dp} dt dp + \int_{\underline{\alpha}}^{q} p \int_{p}^{q'} [1-t] \frac{dQ(F;t)}{dt} \frac{dQ(F;p)}{dp} dt dp. \tag{38}$$

from which (21) follows by writing $p := F(y)$ and $t := F(u)$.

To derive (22) note that, rearranging (21), we have

$$\int_{x_{\underline{\alpha}}}^{x_q} F(y) \int_{y}^{x_{q'}} du dy - \int_{x_{\underline{\alpha}}}^{x_q} F(y) \int_{y}^{x_{q'}} F(u) du dy + \int_{x_{\underline{\alpha}}}^{x_q} \int_{x_{\underline{\alpha}}}^{y} F(u) du dy - \int_{x_{\underline{\alpha}}}^{x_q} F(y) \int_{x_{\underline{\alpha}}}^{y} F(u) du dy \tag{39}$$

which then becomes

$$\begin{aligned}
& \int_{x_{\underline{\alpha}}}^{x_q} F(y) \int_{y}^{x_{q'}} du dy + \int_{x_{\underline{\alpha}}}^{x_q} \int_{x_{\underline{\alpha}}}^{y} F(u) du dy - \int_{x_{\underline{\alpha}}}^{x_q} F(y) \int_{x_{\underline{\alpha}}}^{x_{q'}} F(u) du dy \\
= & \int_{x_{\underline{\alpha}}}^{x_q} F(y) \int_{x_{\underline{\alpha}}}^{x_{q'}} (1-F(u)) du dy - \int_{x_{\underline{\alpha}}}^{x_q} F(y) \int_{x_{\underline{\alpha}}}^{x_{q'}} du dy + \int_{x_{\underline{\alpha}}}^{x_q} F(y) \int_{y}^{x_{q'}} du dy + \int_{x_{\underline{\alpha}}}^{x_q} \int_{x_{\underline{\alpha}}}^{y} F(u) du dy \\
= & \int_{x_{\underline{\alpha}}}^{x_q} F(y) dy \int_{x_{\underline{\alpha}}}^{x_{q'}} (1-F(u)) du + \int_{x_{\underline{\alpha}}}^{x_q} \int_{x_{\underline{\alpha}}}^{y} F(u) du dy - \int_{x_{\underline{\alpha}}}^{x_q} F(y) \int_{x_{\underline{\alpha}}}^{y} du dy \\
= & \int_{x_{\underline{\alpha}}}^{x_q} F(y) dy \int_{x_{\underline{\alpha}}}^{x_{q'}} [1-F(u)] du + \int_{x_{\underline{\alpha}}}^{x_q} \left[[uF(u)]_{x_{\underline{\alpha}}}^{y} - \int_{x_{\underline{\alpha}}}^{y} u dF(u) \right] dy - \int_{x_{\underline{\alpha}}}^{x_q} F(y)[y - x_{\underline{\alpha}}] dy \\
= & \int_{x_{\underline{\alpha}}}^{x_q} F(y) dy \int_{x_{\underline{\alpha}}}^{x_{q'}} (1-F(u)) du + \int_{x_{\underline{\alpha}}}^{x_q} [yF(y) - \underline{\alpha} x_{\underline{\alpha}}] dy -
\end{aligned}$$

$$\int_{x_{\underline{\alpha}}}^{x_q} \int_{x_{\underline{\alpha}}}^{y} u dF(u) dy - \int_{x_{\underline{\alpha}}}^{x_q} F(y) y dy + x_{\underline{\alpha}} \int_{x_{\underline{\alpha}}}^{x_q} F(y) dy \tag{40}$$

from which (22) follows.

A.3 Equivalence of trimmed-variance expresssions

Equations (19) and (27) can be shown to be equivalent. Using

$$\begin{aligned} [1-\alpha]\, c_{\alpha,q} &= c_q - c_{\underline{\alpha}} \\ [1-\alpha]\, s_{\alpha,q} &= s_q - s_{\underline{\alpha}} \end{aligned}$$

expression (27) becomes

$$\begin{aligned} & [q x_q - \underline{\alpha} x_{\underline{\alpha}} - [c_q - c_{\underline{\alpha}}]] \, [[1-q']\, x_{q'} + \underline{\alpha} x_{\underline{\alpha}} + [c_{q'} - c_{\underline{\alpha}}]] \\ & - [x_q \, [c_q - c_{\underline{\alpha}}] - [s_q - s_{\underline{\alpha}}]] - \underline{\alpha} x_{\underline{\alpha}} \, [x_q - x_{\underline{\alpha}}] \\ = \quad & s_q + [q x_q - c_q - [\underline{\alpha} x_{\underline{\alpha}} - c_{\underline{\alpha}}]] \, [[1-q']\, x_{q'} + c_{q'} + \underline{\alpha} x_{\underline{\alpha}} - c_{\underline{\alpha}}] - x_q c_q \\ & + [x_q c_{\underline{\alpha}} - s_{\underline{\alpha}}] - \underline{\alpha} x_{\underline{\alpha}} \, [x_q - x_{\underline{\alpha}}] \\ = \quad & s_q + [q x_q - c_q] \, [[1-q']\, x_{q'} + c_{q'}] - x_q c_q \\ & - [\underline{\alpha} x_{\underline{\alpha}} - c_{\underline{\alpha}}] \, [[1-q']\, x_{q'} + c_{q'} + \underline{\alpha} x_{\underline{\alpha}} - c_{\underline{\alpha}}] + [q x_q - c_q] \, [\underline{\alpha} x_{\underline{\alpha}} - c_{\underline{\alpha}}] \\ & + [x_q c_{\underline{\alpha}} - s_{\underline{\alpha}}] - \underline{\alpha} x_{\underline{\alpha}} \, [x_q - x_{\underline{\alpha}}] \\ = \quad & s_q + [q x_q - c_q] \, [[1-q']\, x_{q'} + c_{q'}] - x_q c_q \\ & + [\underline{\alpha} x_{\underline{\alpha}} - c_{\underline{\alpha}}] \, [q x_q - c_q - [1-q']\, x_{q'} - c_{q'} - \underline{\alpha} x_{\underline{\alpha}} + c_{\underline{\alpha}}] \end{aligned}$$

$$-s_{\underline{\alpha}} + [\underline{\alpha} x_{\underline{\alpha}} - c_{\underline{\alpha}}]\,[x_{\underline{\alpha}} - x_q] + x_{\underline{\alpha}} c_{\underline{\alpha}}$$

which becomes (19) in one step.