# Inference for Lorenz Curves*

Gholamreza Hajargasht and William E. Griffiths

*Department of Economics*
*University of Melbourne, Australia*

January 22, 2016

**ABSTRACT**

The *Lorenz curve*, introduced more than 100 years ago, is still one of the main tools in poverty and inequality analysis. International institutions such as the World Bank collect and publish grouped income data in the form of population and income shares for a large number of countries. These data are often used for estimation of parametric Lorenz curves which in turn form the basis for most poverty and inequality analyses. Despite the prevalence of parametric estimation of Lorenz curves from grouped data, and the existence of well-developed nonparametric methods, a rigorous statistical foundation for estimating parametric Lorenz curves has not been provided. In this paper we propose a sound statistical framework for making inference about parametric Lorenz curves for both grouped and individual data. Building on two data generating mechanisms, efficient methods of estimation and inference are proposed and a number of results useful for comparing the two methods of inference, and aiding computation, are derived. Simulations are used to assess the estimators, and curves are estimated for some example countries. We also show how the proposed methods improve upon World Bank methods and make recommendations for improving current practices.

*Keywords*: Minimum Distance, GMM, GB2 Distribution, General Quadratic, Beta Lorenz Curve, Gini Coefficient, Poverty Measures, Quantile Function Estimation

*JEL Classification*: C13, C16, D31

***Corresponding author:***
William Griffiths
Department of Economics
University of Melbourne
Vic 3010, Australia
Phone: +613 8344 3622
Email for Griffiths: wegrif@unimelb.edu.au
Email for Hajargasht: har@unimelb.edu.au

* This research was supported by Australian Research Council Grant DP140100673. The authors are grateful to Jenny Williams whose comments on an earlier draft improved the exposition.

## 1. Introduction

The Lorenz curve introduced more than 100 years ago (Lorenz 1905) provides an intuitive and complete characterisation of an income distribution and provides a basis for poverty and inequality measurements through, for example, its relation with the Gini coefficient, poverty measures and Lorenz orderings. The modern literature on Lorenz curves sparked by the seminal papers of Atkinson (1970) and Gastwirth (1971) is now substantial both in economics and statistics. Data availability has made the estimation and use of Lorenz curves widespread. For example, institutions such as the *World Bank* and the *World Institute for Development Economics Research* (WIDER) collect and publish income or expenditure data that cover a large number of countries over time. These data are usually in the form of population and income shares for a number of groups, typically between 10 and 20. Parametric Lorenz curves estimated using these grouped data form the basis for poverty and inequality analyses. For example, the World Bank website *PovcalNet* provides a variety of poverty and inequality measures for the countries based on such Lorenz curve estimations.[1]

Despite the Lorenz curve's long history, its importance for welfare analysis, the abundance of parametric Lorenz curves that have been estimated, and the existence of well-developed distribution-free methods of analysis, current practice for parametric Lorenz curve estimation lacks a solid statistical foundation. The main objective of this paper is to remedy this deficiency by providing a sound statistical framework for conducting estimation and inference for parametric Lorenz curves and their by-products. Our paper represents the first description of optimal estimation techniques for both parametric Lorenz curves and related quantile functions, using both grouped and individual data. As a secondary objective we show how the methods we develop lead to an improvement over conventional techniques.

[1] Recently, the World Bank has started to publish data on *PovcalNet* in the form of 100 groups for some countries and years and to base its poverty and inequality estimates on these groups.

Our contribution can be summarised as follows: (*i*) First, focusing on grouped data, we recognise two possible data generation processes and propose two inferential procedures based on minimum distance and generalised method of moments estimation theory. (*ii*) We derive closed forms for the optimal weight matrices for both cases. (*iii*) We uncover the relationship between the two cases, derive expressions that facilitate computation, and show that, under “equivalent conditions”, the two methods provide the same asymptotic covariance matrix for the estimated parameters. (*iv*) Monte Carlo simulations are performed to compare the two methods with each other and with conventional approaches, and to study their finite sample performance. (*v*) We point out weaknesses of procedures used by the World Bank, and show that our methods lead to significant improvements. (*vi*) We also consider estimation with individual data (and a large number of groups) and compare Lorenz curve estimation with quantile and density estimation; we prove that all three approaches provide the same asymptotic covariance matrix for the estimated parameters, widening opportunities for researchers to specify Lorenz or quantile functions as alternatives to density functions.

Our contributions in this paper extend our earlier work [Hajargasht et al. (2012), Griffiths and Hajargasht (2015)] in several important ways. The previous two papers focused solely on generalised method of moments inference for direct representations of income distributions, namely, distribution functions and moment distributions, estimated using grouped data arranged non-cumulatively. Motivated by the special status of Lorenz curves, and their abundance in a long history of publication, this paper develops minimum distance inference for the dual representations – the Lorenz curve and the quantile function – which have been estimated sub-optimally up until now. Other advances in the current paper not appearing in our earlier work are consideration of two data generating processes for grouping the data, the relationship between the two approaches, the development of closed form solutions for weight matrices, applications to show the shortcomings of other works, and an extension to individual as well as grouped data.

The literature on Lorenz curves can be loosely grouped into three categories with the first two frequently overlapping: (*i*) papers that suggest alternative functional forms or families of functional forms for Lorenz curves, (*ii*) papers that focus on parametric estimation of Lorenz curves, and (*iii*) papers that propose distribution free inference for Lorenz curve ordinates. Excellent reviews of the literature on alternative functional forms, their properties, and their relationships with income distributions, can be found in Sarabia (2008) and Kleiber and Kotz (2003). More recent work that falls into the first two categories is that of Wang et al (2011), Wang and Smyth (2015a, 2015b) and Sarabia et al. (2015). Most methods for estimating parametric Lorenz curves have involved linear or nonlinear least squares, or a form of generalised least squares, applied directly to income and population shares, or functions of them. See, for example, Kakwani and Podder (1973, 1976), Basmann et al (1990), Chotikapanich (1993) and Wang et al (2011). For their poverty and inequality analyses, the World Bank has used least squares to estimate both the general quadratic (Villasenor and Arnold, 1989) and the beta (Kakwani, 1980) Lorenz curves, with the better-fitting one being chosen for later analysis.[2] A significant problem with existing approaches to parametric estimation is the lack of a transparent data generating process that is needed for providing a sound basis for inference about the Lorenz curve parameters, and the inequality and poverty measures derived from them. While naïve application of least squares may be a reasonable "fitting device", when used for inference, it does not recognise that observations on cumulative proportions (or their log transformations) are not independent; nor does it utilise the size of the sample that generated the grouped observations. Some attempts have been made to mitigate these concerns. Kakwani and Podder (1976) recognised the

[2] Datt (1998) describes the World Bank's methods. An example of where the estimates are used for poverty and inequality analysis is Chen and Ravallion (2010).

multinomial nature of grouped data and used a Lorenz curve specification that, after transformation, could be placed in an approximate linear model framework. Sarabia et al (1999) suggested a distribution-free method of estimation with bootstrap standard errors, and Chotikapanich and Griffiths (2002) proposed a Dirichlet distributional assumption that is consistent with the proportional nature of the data. Nevertheless, these studies still do not come to grips with the need for a suitable underlying data generation process.

There is also a rich literature [see e.g., Gastwirth (1972), Goldie (1977), Beach and Davidson (1983), Gastwirth and Gail (1985), Bishop et al (1989), Cowell and Victoria-Feser (2002) and references cited therein] that does not make any parametric assumptions about the income distribution or the Lorenz curve. Studies in this literature provide a sound basis for statistical inference for Lorenz ordinates. However, to use these methods with grouped data requires knowledge of the bounds and variances of each group, information that is not typically provided by the World Bank and WIDER. Also, inferences are often made only at observed Lorenz ordinates. Thus, there are compelling reasons for considering a rigorous approach to parametric estimation of Lorenz curves from grouped data. In addition, Ryu and Slottje (1996, 1998) argue that parametric models have merit because of their parsimony.

To tackle the problem of parametric Lorenz curve inference from grouped data, we develop two models based on two data generating processes (DGPs) that correspond to two methods of grouping observations, and propose minimum distance or generalised method of moments (GMM) estimators. In the first method of grouping, the proportion of observations in each group is predetermined prior to sampling, making the group income proportions and the group boundaries random. For this scenario, we use the results of Beach and Davidson (1983) and subsequent literature to specify an optimal weight matrix for a minimum distance estimation procedure which can be used to directly estimate a parametric Lorenz curve, or to estimate a Lorenz curve derived in terms of the parameters of a parametric income distribution. Our approach works without knowing the group boundaries and variances but the cost is to make parametric assumptions about either the income distribution or the Lorenz curve.

In the second method of grouping, the group boundaries are predetermined prior to sampling, making both the proportions of observations and the income proportions in each group random. In this case we build on the work of Hajargasht et al. (2012) and Chotikapanich et al. (2007) to propose a GMM inferential framework for estimating a Lorenz curve that has been derived from a specified parametric income distribution. This framework can also be used to estimate a parametric Lorenz curve providing derivation of the distribution function for income in terms of the Lorenz curve parameters is tractable. Using Monte Carlo and real data examples, we show that our approach performs well for both grouping mechanisms and can improve on conventional estimates in important ways.

Compared to current practice of estimating Lorenz curves, having an inferential procedure built on a sound statistical basis has several advantages. First, it provides the most efficient estimator under specified assumptions. Second, it provides a theoretically sound basis for constructing confidence intervals and conducting hypothesis tests. Third, it has the potential to guide improved methods for grouping or data construction designs, for example, in terms of the number or location of the groups. Fourth, there are features in income data such as censoring, trimming, heavy tails, measurement errors, and survey sampling designs that have important implications for welfare measurement (see e.g., Cowell and Victoria-Feser (2002, 2007, 2008)); theoretically sound frameworks such as the one proposed in this paper have the potential to address these features.

The structure of the paper is as follows. In Section 2 we introduce the notation and describe the two data generating processes. Two examples of Lorenz curves are presented: a parametric Lorenz curve that is specified directly, and one that is derived from a parametric income distribution. In Sections 3 and 4 we consider estimation and inference under each of the two data generation processes, with some of Section 3 also devoted to a comparison of our proposed estimation methods with traditional approaches. The focus is on generalised Lorenz curves; inference for standard Lorenz curves is discussed in an Appendix. The relationship between the estimators for each of the two data generation processes is examined in Section 5. This relationship is useful for simplifying computations, and we are able to show that the two estimators have the same asymptotic variance. In Section 6, using Monte Carlo examples, the finite sample performance of the proposed estimators are studied and compared. Lorenz curves are estimated for some example countries in Section 7; we show how the World Bank's estimates can be improved by using alternative estimation methods and/or alternative Lorenz curve specifications. We also make suggestions for improving current practice more generally. Finally, in Section 8 we study estimation of Lorenz curves with a large number of groups and individual data and compare the proposed techniques to density and quantile estimation. Several appendices provide details and proofs for the propositions in the paper.

## 2. Definitions and the Two Data Generating Processes

Let $y$ denote income, and let $f(y)$, $F(y)$ and $F_1(y)$ be its density, cumulative distribution, and first-moment distribution functions, respectively. We treat the support of $y$ as $[0,\infty)$, but our results hold for a finite support $[y_{\min,}\ y_{\max}]$ providing $f(y)$ is nonzero within this interval. Let $\mu = \int_0^\infty y\, f(y)\, dy$ be the mean of the distribution The Lorenz curve relates the cumulative proportion of income given by

$$\ell = F_1(z) = \frac{1}{\mu}\int_0^z y\, f(y)\, dy \qquad (2.1)$$

to the cumulative proportion of population given by

$$c = F(z) = \int_0^z f(y)\,dy \tag{2.2}$$

It is defined as the set of points $(c,\ell) = \left(F(z), F_1(z)\right)$ created by having *z* range from 0 to $\infty$. Alternatively, using the quantile function $z = F^{-1}(c)$, we can write the Lorenz curve as[3]

$$\ell(c) = \frac{1}{\mu}\int_0^{F^{-1}(c)} y\, f(y)\,dy \qquad 0 < c < 1 \tag{2.3}$$

Since the Lorenz curve is scale invariant, it can only be used for welfare comparisons based solely on income inequality without any consideration for the level or scale of income. The generalised Lorenz curve (GLC), introduced by Shorrocks (1983) to overcome this problem, is the set of points $(c,m) = (c,\mu\ell) = \left(F(z), \mu F_1(z)\right)$ created as *z* ranges from 0 to $\infty$, where

$$m(z) = \mu\ell = \int_0^z y\, f(y)\,dy \tag{2.4}$$

Using the notation *L* to denote the GLC, and replacing $z$ with $F^{-1}(c)$, we have

$$L(c) = \int_0^{F^{-1}(c)} y\, f(y)\,dy \qquad 0 \le c \le 1 \tag{2.5}$$

Here we focus on estimation of the GLC; in an appendix we show how the estimation can be modified to cover the standard Lorenz curve.

In what follows we consider estimation of a vector of parameters $\phi$ which can be the parameters of a parametric Lorenz curve specified directly, or the parameters of a parametric income distribution from which a Lorenz curve has been derived. For the former approach, many Lorenz candidates have been suggested in the literature.[4] Any non-decreasing convex

[3] Lorenz curves can be defined alternatively as $\ell(c) = (1/\mu)\int_0^c F^{-1}(x)\,dx$ (see e.g. Gastwirth 1971). It can be shown that the two definitions coincide if $F(y)$ is continuous (see Iritani and Kuga 1983).

[4] See Sarabia (2008) and Kleiber and Kotz (2003) for reviews, and Wang et al (2011) and Wang and Smyth (2015b) for further proposals.

function $L(c)$ where $c \in [0,1]$, $L(0^+) = 0$, and $L(1^-) = \mu$, can be a generalised Lorenz curve.[5] An example considered later in the paper is the Lorenz curve

$$L(c;\phi) = \mu c^{\beta_1}\left\{1-(1-c)^{\beta_2}\right\}^{\beta_3} \tag{2.6}$$

This curve was proposed by Sarabia et al. (1999) as a generalisation of the models in Rasche et al. (1980) and Ortega et al. (1991). Henceforth we refer to it as the SCS curve in line with the names of its developers. Sufficient conditions for it to satisfy the requirements of a Lorenz curve are $\beta_1 > 0$, $\beta_2 > 0$, $\mu > 0$ and $\beta_3 > 1$. If our objective is to estimate the Lorenz curve in (2.6), then $\phi' = (\beta_1, \beta_2, \beta_3, \mu)$.

As an example of Lorenz curve estimation that begins by specifying a parametric income distribution, consider the generalised beta distribution of the second kind (GB2)

$$f(y;\phi) = \frac{ay^{ap-1}}{b^{ap}\mathrm{B}(p,q)[1+(y/b)^a]^{p+q}} \tag{2.7}$$

with positive parameters $\phi' = (b, p, q, a)$, and with $\mathrm{B}(p,q)$ denoting the beta function. This distribution has been a flexible and popular candidate for income distributions (see e.g., McDonald 1984, McDonald and Ransom 2008, or Kleiber and Kotz 2003). Its cumulative distribution function (cdf) is given by

$$F(y;\phi) = \frac{1}{\mathrm{B}(p,q)}\int_0^u t^{p-1}(1-t)^{q-1}dt = B(u;p,q)$$

where $u = y^a \big/ (b^a + y^a)$, and $B(u;p,q)$ is the cdf for the normalised beta distribution defined on the (0,1) interval. To find the Lorenz curve corresponding to (2.5), we use the following results (see e.g., Kleiber and Kotz 2003, Ch.6)

$$\begin{aligned} c &= B(u;p,q) \\ m &= \frac{b\mathrm{B}(p+1/a, q-1/a)}{\mathrm{B}(p,q)} B(u; p+1/a, q-1/a) \end{aligned} \tag{2.8}$$

[5] See, for example, Iritani and Kuga (1983) and Thistle (1989).

where, in line with (2.2) and (2.4), $u = z^a \big/ \left(b^a + z^a\right)$. From the first equation in (2.8) we have $u = B^{-1}(c; p, q)$. Substituting this expression for $u$ into the second equation yields the Lorenz curve in terms of the parameters of the income distribution

$$L(c;\phi) = \frac{b\,\mathrm{B}\left(p + 1/a, q - 1/a\right)}{\mathrm{B}(p,q)} B\left\{B^{-1}(c; p, q); p + 1/a, q - 1/a\right\} \tag{2.9}$$

To provide a statistically sound method of inference for Lorenz curves based on grouped data, whether it be for a directly specified curve such as (2.6), or an indirectly specified curve like (2.9), we first need to know how the groupings have been made. The nature of the groupings defines the DGP, the model, the distribution theory for estimation, and a choice of suitable notation. As a first step towards some notation that we later modify slightly according to the DGP, assume that a sample of $T$ observations $\left(y_1, y_2, \ldots, y_T\right)$ is randomly drawn from the income distribution, and placed into $N$ income groups defined by the group boundaries $(z_0, z_1), (z_1, z_2), \ldots, (z_{N-1}, z_N)$, where $z_0 = 0$ and $z_N = \infty$. Let $T_i$ be the number of observations and $M_i$ total income in the $i$-th group, and set $c_i = T^{-1}\sum_{j=1}^{i} T_j$ and $\tilde{y}_i = T^{-1}\sum_{j=1}^{i} M_j$. We consider two ways in which the data can be grouped.[6]

*DGP 1: Fixed $c_i$ and stochastic $z_i$*

Here the observations are grouped such that the proportion of observations in each group is pre-specified. Examples are 10 groups with 10% of the observations in each group or 20 groups with 5% of the observations in each group. In this case, the cumulative proportions $c_i$ are fixed (non-random) and the sample group boundaries as well as the average cumulative incomes $\tilde{y}_i$, are random variables. The sample group boundaries are given by $\tilde{z}_i = \max\left\{y_t h_i(y_t)\right\}$, where $h_i(y_t)$ is an indicator function equal to one if $y_t$ is in the $i$-th

[6] In an unpublished work, Wu (2006) has also distinguished between these two cases and has derived estimators for densities with grouped data. Here, our focus is on Lorenz curves and therefore the resulting moment conditions and weight matrix are different.

group and zero otherwise, and where we use a tilde "~" on $\tilde{z}_i$ to recognise its randomness. Define vectors $\tilde{\mathbf{z}}' = (\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_{N-1})$ and $\tilde{\mathbf{y}}'_L = \left(\tilde{y}_{L1}, \tilde{y}_{L2}, \ldots, \tilde{y}_{LN}\right)$. We use the subscript $L$ on $\tilde{\mathbf{y}}_L$ to distinguish it from the average cumulative incomes that we later introduce under *DGP 2*. The sample group boundary $\tilde{z}_i$ is an estimator for the quantile $F^{-1}(c_i;\boldsymbol{\phi})$, and the average cumulative income $\tilde{y}_{Li}$ is an estimator for the Lorenz ordinate $L(c_i;\boldsymbol{\phi})$. If the available grouped data includes information on $\tilde{\mathbf{z}}$ as well as data on $\mathbf{c}' = \left(c_1, c_2, \ldots, c_{N-1}\right)$ and $\tilde{\mathbf{y}}_L$, then both $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{y}}_L$ can be used to estimate $\boldsymbol{\phi}$. If the sample group boundaries are not available, one can proceed with estimation based solely on $\tilde{\mathbf{y}}_L$.[7]

*DGP 2: Stochastic $c_i$ and fixed $z_i$*

The second way in which grouping can take place is with pre-specified group boundaries in which case $\mathbf{z}' = (z_1, z_2, \ldots, z_{N-1})$ is predetermined and the cumulative population shares $c_i$ and average cumulative incomes $\tilde{y}_i$ are random. An example would be $z_1 = \$30,000$, $z_2 = \$60,000$, $z_3 = \$90,000$, and so on, although equal intervals are not essential. In this case we use the notation $\tilde{c}_i = T^{-1}\sum_{j=1}^{i} T_j$ to denote an estimator for $F(z_i;\boldsymbol{\phi})$, with $\tilde{\mathbf{c}}' = \left(\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_{N-1}\right)$. Introducing a subscript $m$ for the cumulative incomes, we have $\tilde{\mathbf{y}}'_m = \left(\tilde{y}_{m1}, \tilde{y}_{m2}, \ldots, \tilde{y}_{mN}\right)$ where $\tilde{y}_{mi}$ is an estimator for $m(z_i;\boldsymbol{\phi})$. In this case both $\tilde{\mathbf{c}}$ and $\tilde{\mathbf{y}}_m$ can be used to estimate $\boldsymbol{\phi}$; if the group boundaries are not provided, then they can be estimated along with $\boldsymbol{\phi}$.

These issues are taken up in the next two sections, with *DGP 1* being considered in Section 3 and *DGP 2* in Section 4.

[7] The possible unavailability of group boundaries is in line with the data on *PovcalNet* where population and income shares and mean income are provided, from which the $c_i$ and $\tilde{y}_i$ can be readily calculated; the $z_i$ traditionally have not been provided, but have recently been posted for some countries.

## 3. Inference for *DGP 1*, fixed $c_i$ and stochastic $z_i$

Let $\mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi})$ be an $(N-1)$ dimensional vector with $F^{-1}(c_i;\boldsymbol{\phi})$ as its *i*-th element, and let $\mathbf{L}(\mathbf{c};\boldsymbol{\phi})$ be an $N$ dimensional vector with $L(c_i;\boldsymbol{\phi})$ as its *i*-th element To estimate $\boldsymbol{\phi}$, we can use $\tilde{\mathbf{z}}$ as an estimator for $\mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi})$, and $\tilde{\mathbf{y}}_L$ as an estimator for $\mathbf{L}(\mathbf{c};\boldsymbol{\phi})$. Two methods for estimating $\boldsymbol{\phi}$ are considered. The first uses information on both the sample group bounds $\tilde{\mathbf{z}}$ and the average cumulative incomes $\tilde{\mathbf{y}}_L$, and the second is relevant when data on $\tilde{\mathbf{z}}$ are not available – the typical scenario in the literature where only $\tilde{\mathbf{y}}_L$ is used. For the moment, we assume that tractable closed form expressions are available for both the quantile function $F^{-1}(c;\boldsymbol{\phi})$ and the Lorenz curve $L(c;\boldsymbol{\phi})$ irrespective of whether $\boldsymbol{\phi}$ contains parameters from a parametric Lorenz curve or a parametric income distribution. We defer discussion of tractability and the relationships between the functions until after the first estimator has been presented.

Assuming that the sample is drawn from a population with cdf $F(y)$, where $f(z_i)=F'(z_i)>0$, and $c_i=F(z_i)$, we have the well-known result (see e.g., Beach and Davidson 1983 and references cited therein)

$$\sqrt{T}\left(\tilde{\mathbf{z}}-\mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi})\right)\rightarrow N(\mathbf{0},\boldsymbol{\Omega}_{L,11}) \tag{3.1}$$

where

$$\boldsymbol{\Omega}_{L,11}=\begin{bmatrix} \frac{c_1(1-c_1)}{f(z_1)^2} & \cdots & \frac{c_1(1-c_{N-1})}{f(z_1)f(z_{N-1})} \\ \vdots & \ddots & \vdots \\ \frac{c_1(1-c_{N-1})}{f(z_1)f(z_{N-1})} & \cdots & \frac{c_{N-1}(1-c_{N-1})}{f(z_{N-1})^2} \end{bmatrix} \tag{3.2}$$

To include $\tilde{\mathbf{y}}_L$ in the estimation procedure, we use results from Beach and Davidson (1983) and Cowell and Victoria-Feser (2002) to obtain

$$\sqrt{T}\left(\tilde{\mathbf{y}}_L - \mathbf{L}(\mathbf{c};\boldsymbol{\phi})\right) \to N(\mathbf{0},\boldsymbol{\Omega}_{L,22}) \tag{3.3}$$

where the elements in $\boldsymbol{\Omega}_{L,22}$ are

$$\left[\boldsymbol{\Omega}_{L,22}\right]_{i,j} = m_i^{(2)} + \left(c_i z_i - L(c_i)\right)\left(z_j - c_j z_j + L(c_j)\right) - z_i L(c_i) \qquad \text{for } i \le j \tag{3.4}$$

with

$$m_i^{(2)} = m^{(2)}(z_i) = \int_0^{z_i} y^2 f(y)dy \tag{3.5}$$

Symmetry is used to establish the elements for $i > j$. To combine $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{y}}_L$ in an estimation procedure, we also need the asymptotic covariance matrix $\boldsymbol{\Omega}_{L,12} = \text{cov}\left(\sqrt{T}\tilde{\mathbf{z}}, \sqrt{T}\tilde{\mathbf{y}}_L\right)$. In Appendix 1 we show that

$$\begin{aligned}
\left[\boldsymbol{\Omega}_{L,12}\right]_{i,j} &= \frac{c_i\left(L(c_j) - z_j c_j + z_j\right) - L(c_i)}{f(z_i)} \qquad \text{for } i \le j \\
\left[\boldsymbol{\Omega}_{L,12}\right]_{i,j} &= \frac{(c_i - 1)\left(L(c_j) - z_j c_j\right)}{f(z_i)} \qquad \text{for } i \ge j
\end{aligned} \tag{3.6}$$

Collecting all these results together, we have

$$\sqrt{T}\begin{bmatrix} \tilde{\mathbf{z}} - \mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi}) \\ \tilde{\mathbf{y}}_L - \mathbf{L}(\mathbf{c};\boldsymbol{\phi}) \end{bmatrix} \to N(\mathbf{0},\boldsymbol{\Omega}_L) \qquad \text{where } \boldsymbol{\Omega}_L = \begin{bmatrix} \boldsymbol{\Omega}_{L,11} & \boldsymbol{\Omega}_{L,12} \\ \boldsymbol{\Omega}'_{L,12} & \boldsymbol{\Omega}_{L,22} \end{bmatrix} \tag{3.7}$$

If we have extra information that can be used to estimate the $m_i^{(2)}$, such as group variances, we can use this information along with $\tilde{z}_i$ and $\tilde{y}_{L,i}$ to consistently estimate $\boldsymbol{\Omega}_{L,22}$, making it possible to find standard errors for the Lorenz ordinates $\tilde{y}_{L,i}$ without making any parametric assumptions. However, group variances are seldom if ever provided with grouped data. This problem can be overcome if one is willing to make a parametric assumption about either the income distribution or the Lorenz curve. Making such an assumption means that $m_i^{(2)}$, $z_i$, $f(z_i)$ and $L(c_i)$ can be expressed in terms of the parameters of the income distribution or the Lorenz curve, and a minimum distance framework can be set up to estimate those parameters.

Introducing a zero subscript to distinguish the underlying parameter vector $\boldsymbol{\phi}_0$ from other possible values of $\boldsymbol{\phi}$, we have the following proposition:

***Proposition 1***: Under certain regularity conditions, $\hat{\boldsymbol{\phi}}_L$ defined by

$$\hat{\boldsymbol{\phi}}_L = \arg\min_{\boldsymbol{\phi}} \mathbf{H}'_L(\boldsymbol{\phi})\tilde{\mathbf{W}}_L\ \mathbf{H}_L(\boldsymbol{\phi}) \tag{3.8}$$

(i) is a consistent estimator of $\boldsymbol{\phi}_0$, and

(ii) is asymptotically normal with

$$\sqrt{T}\left(\hat{\boldsymbol{\phi}}_L - \boldsymbol{\phi}_0\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_L) \quad \text{where} \quad \mathbf{V}_L = \left(\frac{\partial \mathbf{H}'_L}{\partial \boldsymbol{\phi}}\mathbf{W}_L\frac{\partial \mathbf{H}_L}{\partial \boldsymbol{\phi}'}\right)^{-1} \tag{3.9}$$

$\mathbf{W}_L = \boldsymbol{\Omega}_L^{-1}(\boldsymbol{\phi}_0)$, and $\tilde{\mathbf{W}}_L = \boldsymbol{\Omega}_L^{-1}(\tilde{\boldsymbol{\phi}})$, where $\tilde{\boldsymbol{\phi}}$ is any consistent estimator for $\boldsymbol{\phi}_0$, and

$$\mathbf{H}_L(\boldsymbol{\phi}) = \begin{bmatrix} \tilde{\mathbf{z}} - \mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi}) \\ \tilde{\mathbf{y}}_L - \mathbf{L}(\mathbf{c};\boldsymbol{\phi}) \end{bmatrix} \tag{3.10}$$

*Proof:* See Appendix 1 for both the proof and the required regularity conditions.

To implement the estimation procedure defined by (3.8), we need to (a) find expressions in terms of $\boldsymbol{\phi}$ for the quantile function $F^{-1}(c)$, the Lorenz curve $L(c)$, and the weight matrix $\mathbf{W}_L$, (b) suggest a minimisation algorithm, and (c) investigate whether computationally convenient forms for the weight matrix can be obtained. First, suppose we begin with a Lorenz curve specification such as that given in (2.6), then the following lemma can be used to derive the required quantities.

***Lemma 1:*** If $L(c_i;\boldsymbol{\phi})$ satisfies the conditions of a Lorenz curve on $(0,1)$, is twice differentiable with $\partial^2 L(c_i,\boldsymbol{\phi})/\partial c_i^2 > 0$ and $c_i = F(z_i;\boldsymbol{\phi})$, then the following relationships hold

$$(i) \qquad z_i = F^{-1}(c_i;\boldsymbol{\phi}) = \frac{\partial L(c_i;\boldsymbol{\phi})}{\partial c_i} \tag{3.11}$$

$$(ii) \qquad f(z_i;\boldsymbol{\phi}) = \frac{1}{\partial^2 L(c_i;\boldsymbol{\phi})/\partial c_i^2} \tag{3.12}$$

$$(iii) \quad m_i^{(2)}(z_i;\phi) = \int_0^{z_i} y^2 f(y;\phi)dy = \int_0^{c_i} \left( \frac{\partial L(x_i;\phi)}{\partial x_i} \right)^2 dx_i \tag{3.13}$$

*Proof:* (*i*) is a well-known result in the Lorenz curve literature (see e.g. Gastwirth 1971), (*ii*) is also well-known and can obtained by using (*i*) and applying the chain rule, and (*iii*) is obtained by using (*i*) and a change of variable for integration.

As an example, applying (*i*) to the Lorenz curve in (2.6), yields

$$z_i = F^{-1}(c_i;\phi) = \frac{\partial L(c_i;\phi)}{\partial c_i} = \mu \left( c_i^{\beta_1} \{1-(1-c_i)^{\beta_2}\}^{\beta_3} \left\{ \frac{\beta_1}{c_i} + \frac{\beta_2 \beta_3 (1-c_i)^{\beta_2 - 1}}{1-(1-c_i)^{\beta_2}} \right\} \right) \tag{3.14}$$

From (*ii*), further differentiation of this function yields an expression for $f(z_i;\phi)$. A value for $m_i^{(2)}(z_i;\phi)$ can be obtained by numerically integrating the right side of (3.13).

If at the outset we begin by specifying a density function rather than a Lorenz curve, then whether or not estimation via (3.8) is tractable will depend on whether the cdf is invertible, either algebraically or computationally. For the GB2 distribution in (2.7), we can readily derive

$$z_i = F^{-1}(c_i;\phi) = b \left( \frac{B^{-1}(c;p,q)}{1-B^{-1}(c;p,q)} \right)^{1/a} \tag{3.15}$$

and the Lorenz curve is given in (3.8), making estimation tractable. In some other cases, such as a mixture of lognormal distributions, inversion of the cdf is not straightforward, making estimation difficult computationally. The remaining ingredient needed for estimation of the GB2 distribution and its Lorenz curve is the quantity $m_i^{(2)}$ which appears in the weight matrix. It is given by

$$m_i^{(2)}(z_i) = m_i^{(2)} = \frac{\mathrm{B}(p+2/a, q-2/a)\mathrm{B}(p,q)}{\mathrm{B}(p+1/a, q-1/a)^2} B(u_i; p+2/a, q-2/a)$$

For a minimisation algorithm one can use any of several methods for implementing minimum distance estimators including a simple two-step, an iterative two-step or a continuously updating estimator. We employ an iterative two-step estimator where in the first stage we find $\hat{\boldsymbol{\phi}}_{L,1} = \arg\min_{\boldsymbol{\phi}} \mathbf{H}'_L(\boldsymbol{\phi})\boldsymbol{\Omega}_{L,1}^{-1}\mathbf{H}_L(\boldsymbol{\phi})$ with $\boldsymbol{\Omega}_{L,1} = \mathbf{I}$ or some other pre-specified positive definite matrix. Using $\hat{\boldsymbol{\phi}}_{L,1}$ we compute $\boldsymbol{\Omega}_{L,2} = \boldsymbol{\Omega}_L\left(\hat{\boldsymbol{\phi}}_{L,1}\right)$, then, in the second stage, we find $\hat{\boldsymbol{\phi}}_{L,2} = \arg\min_{\boldsymbol{\phi}} \mathbf{H}'_L(\boldsymbol{\phi})\boldsymbol{\Omega}_{L,2}^{-1}\mathbf{H}_L(\boldsymbol{\phi})$. We iterate this process until there is no improvement in the objective function. Having the estimates, the following equation can be used to compute the covariance matrix (and standard errors) for $\hat{\boldsymbol{\phi}}_L$ and functions of them such as the Lorenz curve ordinates or inequality measures

$$\text{var}\left(\hat{\boldsymbol{\phi}}_L\right) = \frac{1}{T}\left(\frac{\partial \hat{\mathbf{H}}'_L}{\partial \boldsymbol{\phi}}\hat{\mathbf{W}}_L\frac{\partial \hat{\mathbf{H}}_L}{\partial \boldsymbol{\phi}'}\right)^{-1} \tag{3.16}$$

Computationally, it is convenient if we can obtain a closed form solution for the inverse $\mathbf{W}_L = \boldsymbol{\Omega}_L^{-1}$, and simplify computations in other ways, particularly when we have a large number of groups. In Section 5 we use results from the setup under *DGP 2* to derive a closed form expression for $\mathbf{W}_L$. In Appendix 2, we prove the following proposition showing that most elements in the weight matrix are zeroes.

***Proposition 2:***

Each block in the weight matrix $\mathbf{W}_L = \begin{bmatrix} \mathbf{W}_{L,11} & \mathbf{W}_{L,12} \\ \mathbf{W}'_{L,12} & \mathbf{W}_{L,22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Omega}_{L,11} & \boldsymbol{\Omega}_{L,12} \\ \boldsymbol{\Omega}'_{L,12} & \boldsymbol{\Omega}_{L,22} \end{bmatrix}^{-1}$ is tri-diagonal.

So far the assumption has been that we have observations on $\tilde{z}_i$. However, there are many important data sets, for example those provided by the World Bank or WIDER, that do not report $\tilde{z}_i$ 's. Fortunately, with parametric assumptions we can still estimate $\boldsymbol{\phi}$ by considering only the second set of equations in (3.10) from which we obtain the estimator

$$\hat{\boldsymbol{\phi}}_{L,0} = \arg\min_{\boldsymbol{\phi}} \left(\tilde{\mathbf{y}}_L - \mathbf{L}(\mathbf{c};\boldsymbol{\phi})\right)' \boldsymbol{\Omega}_{L,22}^{-1}\left(\tilde{\mathbf{y}}_L - \mathbf{L}(\mathbf{c};\boldsymbol{\phi})\right) \tag{3.17}$$

This method of estimation is the closest to the traditional method of estimation of Lorenz curves and for this reason we place special emphasis on it. An estimate of its asymptotic covariance matrix is

$$\operatorname{var}\left(\hat{\boldsymbol{\phi}}_{L,0}\right) = \frac{1}{T}\left(\frac{\partial\hat{\mathbf{L}}'}{\partial\boldsymbol{\phi}}\hat{\boldsymbol{\Omega}}_{L,22}^{-1}\frac{\partial\hat{\mathbf{L}}}{\partial\boldsymbol{\phi}'}\right)^{-1} \tag{3.18}$$

As one would expect, this estimator is asymptotically less efficient than that in equation (3.8) where more information is used (see Appendix 1 for a proof). Other characteristics of (3.17) are that the weight matrix, in this case $\boldsymbol{\Omega}_{L,22}^{-1}$, is no longer tri-diagonal, and it no longer depends on $f(z_i)$; only on $m_i^{(2)}$, $c_i$, $z_i$, and $L(c_i)$.

### *3.1 Comparison with Conventional Estimation Methods*

The estimator in (3.17) provides a context useful for examining most conventional methods for estimating Lorenz curves. Typically, income shares are regressed against a Lorenz function of population shares using linear or nonlinear least squares (e.g., Kakwani 1980, Basmann et al 1990, Chotikapanich 1993, Ryu and Slottje 1996, Datt 1998, and the World Bank website *PovcalNet*). In general, this estimation problem can be written as

$$\min_{\boldsymbol{\phi}} \sum_{i=1}^{N-1}\left\{\tilde{\ell}_i - \ell(c_i;\boldsymbol{\phi})\right\}^2 \tag{3.19}$$

where the $\tilde{\ell}_i$ are observed income shares, the $c_i$ are observed population shares and $\ell(c_i;\boldsymbol{\phi})$ is a parametric Lorenz curve. The generalised Lorenz curve version can be written as[8]

$$\min_{\boldsymbol{\phi}} \sum_{i=1}^{N}\left\{\tilde{y}_i - L(c_i;\boldsymbol{\phi})\right\}^2 \tag{3.20}$$

These methods of estimation ignore the fact that observations on cumulative income proportions are not independent, the consequences of which are inefficient estimates for $\boldsymbol{\phi}$, and incorrect covariance matrices for making inferences about $\boldsymbol{\phi}$, and quantities that depend

[8] In (3.19) $\boldsymbol{\phi}$ does not include $\mu$ whereas it would typically be included in (3.20).

on $\boldsymbol{\phi}$, such as the Gini coefficient[9]. The estimators from (3.20) and (3.17) are identical if $\boldsymbol{\Omega}_{L,22}^{-1} = \mathbf{I}$. Thus, minimising (3.20) yields a consistent estimator for $\boldsymbol{\phi}$ that can be employed (as we do) to find a first-stage estimate for $\boldsymbol{\Omega}_{L,22}^{-1}$ for use in a repeated two-stage estimation framework applied to (3.17). However, ignoring the inefficiency that arises from using the estimates from (3.20) as the end result can make a difference. As is shown in our simulations and examples in Sections 5 and 6, the diagonal elements in $\boldsymbol{\Omega}_{L,22}^{-1}$ are substantially different from each other. In (3.17) the weights applied to the initial elements in $\tilde{\mathbf{y}}_L - \mathbf{L}(\mathbf{c};\boldsymbol{\phi})$ are much larger than the weights applied to the later elements, leading to estimates that can be substantially different from those obtained by minimising (3.20).[10]

The second issue with conventional least-squares based estimation approaches is that they have overlooked using the correct covariance matrix for making inferences about $\boldsymbol{\phi}$. To obtain this covariance matrix, we note that, from minimum distance estimation theory (see e.g., Newey-McFadden 1994, theorems 2.1 and 3.2), and for any positive semi-definite weight matrix $\boldsymbol{\Xi}$, the estimator

$$\hat{\boldsymbol{\phi}} = \arg\min_{\boldsymbol{\phi}} \left(\tilde{\mathbf{y}}_L - \mathbf{L}(\mathbf{c};\boldsymbol{\phi})\right)' \boldsymbol{\Xi} \left(\tilde{\mathbf{y}}_L - \mathbf{L}(\mathbf{c};\boldsymbol{\phi})\right) \tag{3.21}$$

is consistent and asymptotically normal with covariance matrix estimated by

$$\text{var}\left(\hat{\boldsymbol{\phi}}\right) = \frac{1}{T}\left\{ \left( \frac{\partial \hat{\mathbf{L}}'}{\partial \boldsymbol{\phi}} \boldsymbol{\Xi} \frac{\partial \hat{\mathbf{L}}}{\partial \boldsymbol{\phi}'} \right)^{-1} \frac{\partial \hat{\mathbf{L}}'}{\partial \boldsymbol{\phi}} \boldsymbol{\Xi} \boldsymbol{\Omega}_{L,22}(\hat{\boldsymbol{\phi}}) \boldsymbol{\Xi} \frac{\partial \hat{\mathbf{L}}}{\partial \boldsymbol{\phi}'} \left( \frac{\partial \hat{\mathbf{L}}'}{\partial \boldsymbol{\phi}} \boldsymbol{\Xi} \frac{\partial \hat{\mathbf{L}}}{\partial \boldsymbol{\phi}'} \right)^{-1} \right\} \tag{3.22}$$

In the least-squares case where $\boldsymbol{\Xi} = \mathbf{I}$, this reduces to

$$\text{var}\left(\hat{\boldsymbol{\phi}}\right) = \frac{1}{T}\left\{ \left( \frac{\partial \hat{\mathbf{L}}'}{\partial \boldsymbol{\phi}} \frac{\partial \hat{\mathbf{L}}}{\partial \boldsymbol{\phi}'} \right)^{-1} \frac{\partial \hat{\mathbf{L}}'}{\partial \boldsymbol{\phi}} \boldsymbol{\Omega}_{L,22}(\hat{\boldsymbol{\phi}}) \frac{\partial \hat{\mathbf{L}}}{\partial \boldsymbol{\phi}'} \left( \frac{\partial \hat{\mathbf{L}}'}{\partial \boldsymbol{\phi}} \frac{\partial \hat{\mathbf{L}}}{\partial \boldsymbol{\phi}'} \right)^{-1} \right\} \tag{3.23}$$

---

[9] Kakwani and Podder (1976) and Chotikapanich and Griffiths (2002) have proposed ways to deal with these facts but their proposals are not built on transparent data generating processes and hence are not fully satisfactory.

[10] Our discussion has been in terms of the generalised Lorenz curve. Using the results in Appendix 4, it can be easily modified to be applicable to standard Lorenz curves.

This is a more satisfactory covariance matrix than that routinely produced by least squares estimation.

It is also informative to put the "balanced fit" estimation procedure proposed by Wang et al (2011) within the context of the minimum distance estimator in (3.8) where information on both the Lorenz ordinates and the quantile function is used. They recommend minimising

$$b\sum_{i=1}^{N-1}\left\{\tilde{\ell}_i - \ell(c_i;\phi)\right\}^2 + (1-b)\sum_{i=1}^{N-1}\left\{\tilde{z}_i - \mu\frac{\partial\ell(c_i;\phi)}{\partial c_i}\right\}^2$$

where $0 \le b \le 1$ is a pre-specified constant. Within our generalised Lorenz curve framework in (3.8), this approach is equivalent to setting $\mathbf{W}_{L,11} = b\mathbf{I}$, $\mathbf{W}_{L,22} = (1-b)\mathbf{I}$ and $\mathbf{W}_{L,12} = \mathbf{0}$. It is less efficient than using the optimal weight matrix. The relevant covariance matrix for finding standard errors can be found by adapting (3.22) to the case where both sets of information are used.[11]

## 4. Inference for *DGP 2*, stochastic $c_i$ and fixed $z_i$

Now we turn to the second DGP where the group boundaries $\mathbf{z}$ can be viewed as having been set prior to sampling and the random quantities are the cumulative proportions $\tilde{\mathbf{c}}$ and the average cumulative incomes $\tilde{\mathbf{y}}_m$. Let $\mathbf{F}(\mathbf{z};\phi)$ and $\mathbf{m}(\mathbf{z};\phi)$ be $(N-1)$ and $N$-dimensional vectors, respectively, with $i$-th elements $F(z_i;\phi)$ and $m(z_i;\phi)$. Recognising that $\tilde{\mathbf{c}}$ is an estimator for $\mathbf{F}(\mathbf{z};\phi)$ and $\tilde{\mathbf{y}}_m$ is an estimator for $\mathbf{m}(\mathbf{z};\phi)$, we provide the moment conditions, the optimal weight matrix and a GMM strategy for estimating either $\theta' = (\mathbf{z}',\phi')$, or, if the group boundaries are provided, just $\phi$. Let $g_i(y)$ be an indicator function such that $g_i(y) = 1$ if $0 < y \le z_i$ and $g_i(y) = 0$ otherwise, then the cumulative proportions and the average cumulative incomes can be written respectively as

[11] Wang et al (2011) use bootstrapping to obtain standard errors.

$$\tilde{c}_i = \frac{1}{T}\sum_{t=1}^{T} g_i\left(y_t\right) \qquad \tilde{y}_{m,i} = \frac{1}{T}\sum_{t=1}^{T} y_t g_i\left(y_t\right) \tag{4.1}$$

The population moments corresponding to these sample moments are, respectively,

$$F\left(z_i;\phi\right) = E\left[g_i(y)\right] = \int_0^{\infty} g_i(y) f(y) dy = \int_0^{z_i} f(y) dy \tag{4.2}$$

and

$$m\left(z_i;\phi\right) = E\left[y g_i\left(y\right)\right] = \int_0^{\infty} y g_i\left(y\right) f\left(y\right) dy = \int_0^{z_i} y f(y) dy \ . \tag{4.3}$$

Then, the following proposition can be used as a basis for estimation.

***Proposition 3:*** Under *DGP 2* and assuming that the sample is randomly drawn from a distribution with cdf $F(y)$, with finite mean and variance, then

$$\sqrt{T}\begin{bmatrix} \tilde{\mathbf{c}} - \mathbf{F}(\mathbf{z};\phi) \\ \tilde{\mathbf{y}}_m - \mathbf{m}(\mathbf{z};\phi) \end{bmatrix} \to N(\mathbf{0}, \mathbf{\Omega}_m)$$

where

$$\mathbf{\Omega}_m = \left[\begin{array}{ccc|ccc} F_1(1-F_1) & \cdots & F_1(1-F_{N-1}) & m_1 - m_1 F_1 & \cdots & m_1 - F_1 m_N \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ F_1(1-F_{N-1}) & \cdots & F_{N-1}(1-F_{N-1}) & m_1 - m_1 F_{N-1} & \cdots & m_{N-1} - F_{N-1} m_N \\ \hline m_1 - m_1 F_1 & \cdots & m_1 - m_1 F_{N-1} & m_1^{(2)} - m_1^2 & \cdots & m_1^{(2)} - m_1 m_N \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ m_1 - F_1 m_N & \cdots & m_{N-1} - F_{N-1} m_N & m_1^{(2)} - m_1 m_N & \cdots & m_N^{(2)} - m_N^2 \end{array}\right] \tag{4.4}$$

and we have used the shorthand notation $m_i = m\left(z_i;\phi\right)$ and $F_i = F\left(z_i;\phi\right)$.

*Proof*: See Appendix 2.

Like before, if in addition to the population shares $\tilde{c}_i$ and income data $\tilde{y}_{m,i}$, the group variances are available, we can consistently estimate standard errors for the $\tilde{c}_i$ and $\tilde{y}_{m,i}$ without making any parametric assumptions. In the absence of group variances, we can proceed by making parametric assumptions. Suppose first that the group boundaries are unknown and are treated as parameters to be estimated. Let the complete set of unknown parameters be given by $\theta_0' = (\mathbf{z}', \phi_0')$.

***Proposition 4:*** Under certain regularity conditions, the GMM estimator $\hat{\boldsymbol{\theta}}$ given by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathbf{H}_m'(\boldsymbol{\theta})\,\tilde{\mathbf{W}}_m\,\mathbf{H}_m(\boldsymbol{\theta}) \tag{4.5}$$

*(i)* is a consistent estimator of $\theta_0$, and

*(ii)* is asymptotically normal with

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_m) \quad \text{where} \quad \mathbf{V}_m = \left(\frac{\partial \mathbf{H}_m'}{\partial \boldsymbol{\theta}} \mathbf{W}_m \frac{\partial \mathbf{H}_m}{\partial \boldsymbol{\theta}'}\right)^{-1} \tag{4.6}$$

$$\mathbf{H}_m(\boldsymbol{\theta}) = \begin{bmatrix} \tilde{\mathbf{c}} - \mathbf{F}(\boldsymbol{\theta}) \\ \tilde{\mathbf{y}}_m - \mathbf{m}(\boldsymbol{\theta}) \end{bmatrix} \tag{4.7}$$

$\mathbf{W}_m = \boldsymbol{\Omega}_m^{-1}(\boldsymbol{\theta}_0)$, $\tilde{\mathbf{W}}_m = \boldsymbol{\Omega}_m^{-1}(\tilde{\boldsymbol{\theta}})$ where $\tilde{\boldsymbol{\theta}}$ is any consistent estimator for $\boldsymbol{\theta}_0$, and $\boldsymbol{\Omega}_m$ is the covariance matrix of the limiting distribution of $\sqrt{T}\,\mathbf{H}_m$ from (4.4).

*Proof:* See Appendix 2 for the proof and the regularity conditions.

If the starting point is specification of a parametric income distribution, then the estimator in (4.5) will be tractable providing we have a satisfactory way for computing values of its cdf and its first and 2nd moment distribution functions. That is, $F_i$, $m_i$ and $m_i^{(2)}$ can be readily calculated. Such is the case for the GB2 distribution, for example. Whether or not estimation is tractable when the starting point is a parametric Lorenz curve will depend on whether the cdf can be obtained by inverting the quantile function. The quantile function given in (3.14) for the Sarabia GLC cannot be readily inverted such that $c_i$ is written as a function of $z_i$, making estimation intractable for this case. More light will be shed on this issue with further examples in Section 6.

Again, an iterative two-step GMM estimator works well in practice. In the first stage, we set $\boldsymbol{\Omega}_{m,1} = \text{Diag}(\tilde{c}_1^{-2},\ldots,\tilde{c}_N^{-2}, \tilde{y}_{m,1}^{-2},\ldots,\tilde{y}_{m,N}^{-2})$ and find $\hat{\boldsymbol{\theta}}_1 = \arg\min_{\boldsymbol{\theta}} \mathbf{H}_m'(\boldsymbol{\theta})\boldsymbol{\Omega}_{m,1}^{-1}\,\mathbf{H}_m(\boldsymbol{\theta})$. In the second stage we find $\hat{\boldsymbol{\theta}}_2 = \arg\min_{\boldsymbol{\theta}} \mathbf{H}'_m\left(\boldsymbol{\theta}\right)\boldsymbol{\Omega}_m^{-1}\left(\hat{\boldsymbol{\theta}}_1\right)\mathbf{H}_m\left(\boldsymbol{\theta}\right)$, and then we iterate until convergence. The rationale behind using $\boldsymbol{\Omega}_{m,1}$ in the first step is that it leads to an estimator that minimizes the sum of squares of the percentage errors in the moment conditions. After estimation, (4.6) can be used to compute standard errors for the elements in $\hat{\theta}$ and functions of them. Note also that when the $z_i$'s are known, the same procedure can be followed, the only difference is that the number of parameters to be estimated is less.

Computations can be simplified by recognising that the 4 blocks in the weight matrix $\mathbf{W}_m$ are tri-diagonal. Commensurate with the partitioning of $\mathbf{H}_m\left(\boldsymbol{\theta}\right)$ in (4.7), let $\mathbf{W}_m$ be partitioned as

$$\mathbf{W}_m = \begin{bmatrix} \mathbf{W}_{m,11} & \mathbf{W}_{m,12} \\ \mathbf{W}'_{m,12} & \mathbf{W}_{m,22} \end{bmatrix}$$

***Proposition 5:*** Let $\kappa_i = \kappa_i\left(\boldsymbol{\theta}\right) = F(z_i;\boldsymbol{\phi}) - F(z_{i-1};\boldsymbol{\phi})$, $\mu_i = \mu_i\left(\boldsymbol{\theta}\right) = m(z_i;\boldsymbol{\phi}) - m(z_{i-1};\boldsymbol{\phi})$, $\mu_i^{(2)} = \mu_i^{(2)}\left(\boldsymbol{\theta}\right) = m^{(2)}(z_i;\boldsymbol{\phi}) - m^{(2)}(z_{i-1};\boldsymbol{\phi})$, and $v_i = \kappa_i \mu_i^{(2)} - \mu_i^2$. Each of the blocks $\mathbf{W}_{m,ij}$ is tri-diagonal, with the following elements

$$\mathbf{W}_{m,11} = \begin{bmatrix} \dfrac{\mu_2^{(2)}}{v_2} + \dfrac{\mu_1^{(2)}}{v_1} & -\mu_2^{(2)}/v_2 & & \\ -\mu_2^{(2)}/v_2 & \ddots & \ddots & \\ & \ddots & \ddots & -\mu_{N-1}^{(2)}/v_{N-1} \\ & & -\mu_{N-1}^{(2)}/v_{N-1} & \dfrac{\mu_N^{(2)}}{v_N} + \dfrac{\mu_{N-1}^{(2)}}{v_{N-1}} \end{bmatrix}$$

$$\mathbf{W}_{m,12} = \begin{bmatrix} \dfrac{-\mu_2}{v_2} - \dfrac{\mu_1}{v_1} & \mu_2/v_2 & & & \\ \mu_2/v_2 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \mu_{N-1}/v_{N-1} & \dfrac{-\mu_N}{v_N} - \dfrac{\mu_{N-1}}{v_{N-1}} & \mu_N/v_N \end{bmatrix}$$

$$\mathbf{W}_{m,22} = \begin{bmatrix} \frac{\kappa_1}{v_1} + \frac{\kappa_2}{v_2} & -\kappa_1/v_1 & & \\ -\kappa_1/v_1 & \ddots & \ddots & \\ & \ddots & \frac{\kappa_{N-1}}{v_{N-1}} + \frac{\kappa_N}{v_N} & -\kappa_{N-1}/v_{N-1} \\ & & -\kappa_{N-1}/v_{N-1} & -\kappa_N/v_N \end{bmatrix}$$

*Proof:* See Appendix 2.

## 5. Some Relationships Between *DGP 1* and *DGP 2*

In terms of implementation**,** two advantages of *DGP 2* are that we have a closed form expression for the weight matrix and we have access to a richer collection of functions with good properties [e.g., generalised beta, mixture of lognormals, etc.] to choose from. However, the model developed for *DGP 1* is more in line with traditional Lorenz curve estimation and, in the absence of information on the class bounds, it can work with only estimating the second (i.e., "mean") set of equations that uses information on $\tilde{\mathbf{y}}_L$. In this Section we give expressions for the relationship between the weight matrices for the two estimation procedures, expressions that can be utilised to compute the weight matrices for estimation under *DGP 1* from the weight matrix under *DGP 2*. Let $\hat{\boldsymbol{\theta}}$ from *DGP* 2 be partitioned as $\hat{\boldsymbol{\theta}}' = \left(\hat{\mathbf{z}}', \hat{\boldsymbol{\phi}}_m\right)$. We also show, under "equivalent conditions" stated in Proposition 6, that the asymptotic variances for the two estimators $\hat{\boldsymbol{\phi}}_L$ and $\hat{\boldsymbol{\phi}}_m$ are the same.

In what follows, we use the following notation: if $\mathbf{a}' = (a_1, \ldots, a_n)$ and $\mathbf{b}' = (b_1, \ldots, b_n)$, then $\mathbf{a.b} = (a_1 b_1, \ldots, a_n b_n)'$, $\mathbf{a}/\mathbf{b} = (a_1/b_1, \ldots, a_n/b_n)'$, and $D[\mathbf{a}]$ is a diagonal matrix with the elements of $\mathbf{a}$ on the diagonal. Also, we use $\mathbf{j}' = (1, \ldots, 1)$ to denote a vector of ones.

***Proposition 6*:** Suppose group bounds $\mathbf{z}$ in *DGP 2* and population shares $\mathbf{c}$ in *DGP 1* are chosen in a way that $\mathbf{z} = \mathbf{F}^{-1}(\mathbf{c})$ then[12]

$$\textit{(i)}\ \boldsymbol{\Omega}_L = \mathbf{A}\boldsymbol{\Omega}_m\mathbf{A}' \qquad \text{with} \quad \mathbf{A} = \begin{pmatrix} D[-\mathbf{j}/\mathbf{f}(\mathbf{z})]_{N-1} & \mathbf{0}_{(N-1)\times N} \\ \begin{bmatrix} -D[\mathbf{z}]_{N-1} \\ \mathbf{0}_{1\times(N-1)} \end{bmatrix} & \mathbf{I}_N \end{pmatrix}$$

$$\textit{(ii)}\ \mathbf{W}_L = \mathbf{A}'^{-1}\mathbf{W}_m\mathbf{A}^{-1} \qquad \text{with} \quad \mathbf{A}^{-1} = \begin{pmatrix} D[-\mathbf{f}(\mathbf{z})]_{N-1} & \mathbf{0}_{(N-1)\times N} \\ \begin{bmatrix} D[-\mathbf{f}(\mathbf{z}).\mathbf{z}]_{N-1} \\ \mathbf{0}_{1\times(N-1)} \end{bmatrix} & \mathbf{I}_N \end{pmatrix}$$

and the lower diagonal blocks of $\mathbf{W}_L$ and $\mathbf{W}_m$ are equal.

$$\textit{(iii)}\ \boldsymbol{\Omega}_{L,22}^{-1} = (\mathbf{B}\boldsymbol{\Omega}_m\mathbf{B}')^{-1} = (\mathbf{B}\mathbf{B}')^{-1}\mathbf{B}[\mathbf{W}_m - \mathbf{W}_m\mathbf{C}(\mathbf{C}'\mathbf{W}_m\mathbf{C})^{-1}\mathbf{C}'\mathbf{W}_m]\mathbf{B}'(\mathbf{B}\mathbf{B}')^{-1}$$

$$\text{with} \quad \mathbf{B} = \begin{pmatrix} \begin{bmatrix} -D[\mathbf{z}]_{N-1} \\ \mathbf{0}_{1\times(N-1)} \end{bmatrix} & \mathbf{I}_N \end{pmatrix} \quad \text{and} \quad \mathbf{C}' = [\mathbf{I}_{N-1} \quad D[\mathbf{z}]_{N-1} \quad \mathbf{0}_{(N-1)\times 1}]$$

*Proof*: *(i)* can be shown by direct multiplication after recognising that $c_i = F(z_i)$ and that this implies $m(z_i) = L(c_i)$. To prove *(ii)* note that $\mathbf{W}_L = \boldsymbol{\Omega}_L^{-1} = \left(\mathbf{A}\boldsymbol{\Omega}_m\mathbf{A}'\right)^{-1} = \mathbf{A}'^{-1}\mathbf{W}_m\mathbf{A}^{-1}$. The particular form of $\mathbf{A}^{-1}$ can be checked by showing that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. The equality of the lower diagonal blocks of $\mathbf{W}_L$ and $\mathbf{W}_m$ can be checked by multiplication. The first equality in *(iii)* comes from (*i*). To prove the second equality in *(iii)*, we first note that

$$\mathbf{B}'(\mathbf{B}\mathbf{B}')^{-1}\mathbf{B} = \mathbf{I} - \mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'$$

from which it follows that

$$[\mathbf{I} - \mathbf{C}(\mathbf{C}'\mathbf{W}_m\mathbf{C})^{-1}\mathbf{C}'\mathbf{W}_m]\,\mathbf{B}'(\mathbf{B}\mathbf{B}')^{-1}\mathbf{B} = [\mathbf{I} - \mathbf{C}(\mathbf{C}'\mathbf{W}_m\mathbf{C})^{-1}\mathbf{C}'\mathbf{W}_m]$$

Then, postmultiplying the right hand side of *(iii)* by $\mathbf{B}\boldsymbol{\Omega}_m\mathbf{B}'$ yields

[12] The result in *(i)* might also be of interest to distribution free studies since the lower diagonal block provides the relationship between the variances of the Lorenz ordinates under the two DGPs, $\tilde{\mathbf{y}}_L$ and $\tilde{\mathbf{y}}_m$.

$$(\mathbf{BB}')^{-1}\mathbf{BW}_m[\mathbf{I}-\mathbf{C}(\mathbf{C}'\mathbf{W}_m\mathbf{C})^{-1}\mathbf{C}'\mathbf{W}_m]\mathbf{B}'(\mathbf{BB}')^{-1}\times[\mathbf{B}\boldsymbol{\Omega}_m\mathbf{B}']$$

$$=(\mathbf{BB}')^{-1}\mathbf{BW}_m[\mathbf{I}-\mathbf{C}(\mathbf{C}'\mathbf{W}_m\mathbf{C})^{-1}\mathbf{C}'\mathbf{W}_m]\boldsymbol{\Omega}_m\mathbf{B}'$$

$$=(\mathbf{BB}')^{-1}\mathbf{BW}_m[\boldsymbol{\Omega}_m\mathbf{B}'-\mathbf{C}(\mathbf{C}'\mathbf{W}_m\mathbf{C})^{-1}\mathbf{C}'\mathbf{B}'] \qquad \text{since } \mathbf{W}_m\boldsymbol{\Omega}_m=\mathbf{I}$$

$$=(\mathbf{BB}')^{-1}\mathbf{BB}' \qquad \text{since } \mathbf{C}'\mathbf{B}'=\mathbf{0}$$

$$=\mathbf{I}$$

Result *(ii)* in this proposition provides a closed form for the weight matrix for *DGP 1* when both quantile and Lorenz equations are considered. The result in (*iii*) provides a tractable formula for the weight matrix when only the Lorenz equations are considered (see equation (3.17)). Its tractability comes from the facts that $\mathbf{BB}'$ is diagonal, $\mathbf{C}'\mathbf{W}_m\mathbf{C}$ is tri-diagonal and there are special formulae for inverting such matrices.

It can also be shown that the two methods provide the same asymptotic variance for $\hat{\boldsymbol{\phi}}$ if the *a priori* groupings from both methods are equivalent. For example, if data are grouped under *DGP 1* into say 10 equal groups with $\mathbf{c}=(0.1,...,0.9)'$ and under *DGP 2* the bounds are specified as $\mathbf{z}=[F(0.1,\boldsymbol{\phi}_0),....,F(0.9,\boldsymbol{\phi}_0)]'$, then the asymptotic variances of the estimated parameters from the two cases are equal[13]. More formally:

***Proposition-7***

*(i)* *Known group bounds*: If $z_i=F^{-1}(c_i;\boldsymbol{\phi}_0)$, then the asymptotic variance of $\hat{\boldsymbol{\phi}}_L$ under *DGP 1* with observed $\{c_i,\tilde{y}_{L,i},\tilde{z}_i\}$ is equal to asymptotic variance of $\hat{\boldsymbol{\phi}}_m$ under *DGP 2* with observed $\{\tilde{c}_i,\tilde{y}_{m,i},z_i\}$. That is,

$$\frac{1}{T}\left(\begin{bmatrix}\dfrac{\partial\mathbf{F}'^{-1}}{\partial\boldsymbol{\phi}} & \dfrac{\partial\mathbf{L}'}{\partial\boldsymbol{\phi}}\end{bmatrix}\mathbf{W}_L(\boldsymbol{\phi})\begin{bmatrix}\partial\mathbf{F}^{-1}/\partial\boldsymbol{\phi}'\\ \partial\mathbf{L}/\partial\boldsymbol{\phi}'\end{bmatrix}\right)^{-1}=\frac{1}{T}\left(\begin{bmatrix}\dfrac{\partial\mathbf{F}'}{\partial\boldsymbol{\phi}} & \dfrac{\partial\mathbf{m}'}{\partial\boldsymbol{\phi}}\end{bmatrix}\mathbf{W}_m(\boldsymbol{\phi})\begin{bmatrix}\partial\mathbf{F}/\partial\boldsymbol{\phi}'\\ \partial\mathbf{m}/\partial\boldsymbol{\phi}'\end{bmatrix}\right)^{-1}$$

where both are evaluated at $\boldsymbol{\phi}_0$.

[13] Note that, after sampling, the resulting grouped data will be different in the two cases even if the grouping is made on the same random sample.

*(ii)* *Unobserved group bounds*: The asymptotic variance of $\hat{\boldsymbol{\phi}}_L$ under *DGP 1* with observed $\{c_i, \tilde{y}_i\}$ is equal to the asymptotic variance of $\hat{\boldsymbol{\phi}}_m$ under *DGP 2* with observed $\{\tilde{c}_i, \tilde{y}_i\}$. That is, $\frac{1}{T}\left(\frac{\partial \mathbf{L}'}{\partial \boldsymbol{\phi}} \boldsymbol{\Omega}_{L,22}^{-1}(\boldsymbol{\phi}) \frac{\partial \mathbf{L}}{\partial \boldsymbol{\phi}'}\right)^{-1}$ evaluated at $\boldsymbol{\phi}_0$ is equal to the lower diagonal block of $\text{cov}\left(\hat{\boldsymbol{\theta}}\right) = \frac{1}{T}\left(\begin{bmatrix} \frac{\partial \mathbf{F}'}{\partial \boldsymbol{\theta}} & \frac{\partial \mathbf{m}'}{\partial \boldsymbol{\theta}} \end{bmatrix} \mathbf{W}_m(\boldsymbol{\theta}) \begin{bmatrix} \partial \mathbf{F}/\partial \boldsymbol{\theta}' \\ \partial \mathbf{m}/\partial \boldsymbol{\theta}' \end{bmatrix}\right)^{-1}$ evaluated at $\boldsymbol{\theta}_0' = [\mathbf{F}^{-1}(\mathbf{c}, \boldsymbol{\phi}_0)', \boldsymbol{\phi}_0']$.

*Proof*: See Appendix 3.

To prove the above proposition we need the following lemma which might be of interest on its own.

***Lemma 2:*** If $c_i = F(z_i; \boldsymbol{\phi})$, then under suitable differentiability conditions, we have the following results:

*(i)* $$\frac{\partial F(z_i;\boldsymbol{\phi})}{\partial \phi_j} = -\frac{\partial F^{-1}(c_i;\boldsymbol{\phi})}{\partial \phi_j} \Bigg/ \frac{\partial F^{-1}(c_i;\boldsymbol{\phi})}{\partial c_i} = -f(z_i;\boldsymbol{\phi}) \frac{\partial F^{-1}(c_i;\boldsymbol{\phi})}{\partial \phi_j} = -\frac{\partial^2 L(c_i;\boldsymbol{\phi})}{\partial c_i \partial \phi_j} \Bigg/ \frac{\partial^2 L(c_i;\boldsymbol{\phi})}{\partial c_i^2}$$

*(ii)* $$\frac{\partial m(z_i;\boldsymbol{\phi})}{\partial \phi_j} = \frac{\partial L(c_i;\boldsymbol{\phi})}{\partial \phi_j} - \frac{\partial L(c_i;\boldsymbol{\phi})}{\partial c_i} \frac{\partial^2 L(c_i;\boldsymbol{\phi})/\partial c_i \partial \phi_j}{\partial^2 L(c_i;\boldsymbol{\phi})/\partial c_i^2}$$

*(iii)* $$\frac{\partial L(c_i;\boldsymbol{\phi})}{\partial \phi_j} = \frac{\partial m(z_i;\boldsymbol{\phi})}{\partial \phi_j} - z_i \frac{\partial F(z_i;\boldsymbol{\phi})}{\partial \phi_j}$$

*Proof*: *(i)* can be obtained by differentiating $F\left(F^{-1}(c_i;\boldsymbol{\phi});\boldsymbol{\phi}\right) = c_i$ with respect to $\phi_j$, applying the chain rule and the results from Lemma 1; *(ii)* is obtained by differentiating $m(z_i;\boldsymbol{\phi}) = L\left(F(z_i,\boldsymbol{\phi});\boldsymbol{\phi}\right)$ with respect to $\phi_j$, applying the chain rule and the result from *(i)*; and *(iii)* is obtained by differentiating $L(c_i;\boldsymbol{\phi}) = m\left(F^{-1}(c_i;\boldsymbol{\phi});\boldsymbol{\phi}\right)$ with respect to $\phi_j$, applying the chain rule and the result from *(i)*.

## 6. Monte Carlo Experiment

Here we perform a Monte Carlo experiment to address the following questions: *(i)* Do the estimators perform well? *(ii)* Does the method of grouping matter? *(iii)* How do the various estimators compare with each other and with some traditional Lorenz curve estimation methods? (*iv*) How do the estimators perform in finite samples?

With the exception of one experiment, the data for all simulations were obtained by drawing 10,000 observations from a GB2 distribution with $b=100$, $p=1$, $q=1.5$ and $a=1.5$, implying a relatively heavy-tailed Singh-Maddala distribution with a Gini coefficient of 0.53.[14] A heavy-tailed case was considered because inference in such cases can be more challenging. Although the data were generated from the Singh-Maddala distribution, the more general GB2 distribution was estimated.

In the first experiment we investigate the performance of covariance matrix estimators for the Lorenz ordinates without making any distributional assumptions. The generated 10,000 observations were used to create 5 groups based on the two grouping methods. For *DGP 1*, the data were grouped by dividing the simulated sample into 5 equally-sized groups with $\mathbf{c}'=(0.2,0.4,\ldots,0.8)$; then the associated $\tilde{y}_i$s and $\tilde{z}_i$s were computed. For *DGP 2*, the groups bounds ($z_i$'s) were set according to the theoretical quantiles corresponding to $\mathbf{c}$, and the associated $\tilde{y}_i$ and $\tilde{c}_i$ were computed. We restricted the number of groups to 5 in this case for ease of presentation of the results. For later experiments we used 20 groups with corresponding cumulative proportions $\mathbf{c}'=(0.05,0.1,\ldots,0.95)$. The two covariance matrices being considered are $\mathbf{\Omega}_{L,22}$ and $\mathbf{\Omega}_{m,22}$ given in (3.4) and (4.4), respectively. In (3.4), $\tilde{y}_{L,i}$ is used to estimate $L(c_i)$; when the grouping is made, we also compute $\tilde{y}_{L,i}^{(2)}=T^{-1}\sum_{t=1}^{T}y_t^2 g_i(y_t)$

[14] A Singh-Maddala distribution is a special case of the GB2 where $p=1$. A necessary and sufficient condition for the existence of the *k*-th moment of the GB2 is $aq>k$. Thus, for $a=q=1.5$, moments beyond the second one do not exist.

which is used to estimate $m_i^{(2)}$. In (4.4), $\tilde{y}_{m,i}$ is used to estimate $m_i$ and $\tilde{y}_{m,i}^{(2)}$ is used to estimate $m_i^{(2)}$.

Table 1 contains the results from 500 replications. Its first panel provides the Monte Carlo averages of the estimates for $\mathbf{\Omega}_{L,22}$ and $\mathbf{\Omega}_{m,22}$. The second panel contains the Monte Carlo covariance matrix of the simulated Lorenz ordinates. The matrices $\mathbf{\Omega}_{L,22}$ and $\mathbf{\Omega}_{m,22}$ evaluated at the true values of the parameters appear in the 3rd panel, and the inverses of these matrices are in the final panel. The values in the first three panels for each of the DGPs are largely similar confirming the validity of the formulae. There are somewhat large differences between alternative variances of the last group mean where $\tilde{y}_5$ is an estimate of the mean $\mu$. This appears to be attributable to the heavy-tailed distribution which is close to having an infinite second order moment. The differences are reduced when the number of Monte Carlo replications is increased and substantially reduced if the data are generated from a distribution with a narrower tail. A comparison of the *DGP 1* and *DGP 2* covariance matrices shows that their variances can be substantially different. The variances for the initial groups are smaller for *DGP 1*, they become larger for the later groups and become equal as expected for the last group where in both cases the mean is being estimated. These results are evident from the formula

$$\text{DGP 1:} \quad \left[\mathbf{\Omega}_{L,22}\right]_{ii} = m_i^{(2)} - m_i^2 + z_i(1-c_i)(c_i z_i - 2m_i)$$

$$\text{DGP 2:} \quad \left[\mathbf{\Omega}_{m,22}\right]_{ii} = m_i^{(2)} - m_i^2$$

As long as $c_i z_i < 2m_i$ (which happens for the initial groups with typical income distributions) the *DGP 1* variance is smaller. Also the variances are equal when $c_i = 1$. We can also see (both theoretically and from the simulation) that the covariances from *DGP 1* are always positive but for *DGP 2* they could be positive or negative. Finally, the inverses of the covariance matrices, i.e., the weight matrices for both DGPs, are starkly different from the identity matrix which is often used implicitly in least square estimation.

Table-1: Covariance matrix for $\tilde{y}_i$ s using a distribution free approach

| | **Covariance Matrix of Lorenz Ordinates *DGP 1*** | | | | | **Covariance Matrix of Lorenz Ordinates *DGP 2*** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Formula (3.4) at Sample Information | | | | | Formula (4.4) at Sample Information | | | | |
| | $\tilde{y}_1$ | $\tilde{y}_2$ | $\tilde{y}_3$ | $\tilde{y}_4$ | $\tilde{y}_5$ | $\tilde{y}_1$ | $\tilde{y}_2$ | $\tilde{y}_3$ | $\tilde{y}_4$ | $\tilde{y}_5$ |
| $\tilde{y}_1$ | 0.377 | 0.822 | 1.249 | 1.722 | 2.598 | 0.578 | 0.296 | -0.182 | -0.973 | -3.193 |
| $\tilde{y}_2$ | 0.822 | 2.405 | 4.130 | 6.037 | 9.573 | 0.296 | 2.912 | 1.253 | -1.496 | -9.203 |
| $\tilde{y}_3$ | 1.249 | 4.130 | 8.419 | 13.589 | 23.179 | -0.182 | 1.253 | 7.793 | 1.731 | -15.267 |
| $\tilde{y}_4$ | 1.722 | 6.037 | 13.589 | 26.117 | 51.969 | -0.973 | -1.496 | 1.731 | 18.314 | -14.072 |
| $\tilde{y}_5$ | 2.60 | 9.57 | 23.18 | 51.97 | 512.75 | -3.19 | -9.20 | -15.27 | -14.07 | 512.75 |
| | Covariance of Simulated Draws | | | | | Covariance of Simulated Draws | | | | |
| $\tilde{y}_1$ | 0.382 | 0.821 | 1.248 | 1.691 | 2.252 | 0.597 | 0.293 | -0.181 | -1.009 | -3.192 |
| $\tilde{y}_2$ | 0.821 | 2.349 | 3.993 | 5.766 | 8.828 | 0.293 | 2.889 | 1.091 | -1.589 | -8.402 |
| $\tilde{y}_3$ | 1.248 | 3.993 | 8.011 | 12.813 | 20.382 | -0.181 | 1.091 | 7.862 | 2.035 | -11.435 |
| $\tilde{y}_4$ | 1.691 | 5.766 | 12.813 | 24.883 | 45.378 | -1.009 | -1.589 | 2.035 | 18.514 | -12.559 |
| $\tilde{y}_5$ | 2.25 | 8.83 | 20.38 | 45.38 | 444.39 | -3.19 | -8.40 | -11.44 | -12.56 | 444.39 |
| | Formula (3.4) Computed at True Values | | | | | Formula (4.4) Computed at True Values | | | | |
| $\tilde{y}_1$ | 0.377 | 0.823 | 1.251 | 1.724 | 2.601 | 0.578 | 0.296 | -0.182 | -0.973 | -3.191 |
| $\tilde{y}_2$ | 0.823 | 2.404 | 4.129 | 6.034 | 9.572 | 0.296 | 2.914 | 1.253 | -1.497 | -9.203 |
| $\tilde{y}_3$ | 1.251 | 4.129 | 8.413 | 13.578 | 23.168 | -0.182 | 1.253 | 7.795 | 1.730 | -15.268 |
| $\tilde{y}_4$ | 1.724 | 6.034 | 13.578 | 26.089 | 51.938 | -0.973 | -1.497 | 1.730 | 18.317 | -14.068 |
| $\tilde{y}_5$ | 2.60 | 9.57 | 23.17 | 51.94 | 615.62 | -3.19 | -9.20 | -15.27 | -14.07 | 615.62 |
| | Weight Matrix Computed at True Values | | | | | Weight Matrix Computed at True Values | | | | |
| $\tilde{y}_1$ | 1335.50 | -777.03 | 220.01 | -23.24 | 0.119 | 207.01 | -15.09 | 7.51 | 10.03 | 1.26 |
| $\tilde{y}_2$ | -777.030 | 806.520 | -389.100 | 67.989 | -0.349 | -15.091 | 42.225 | -7.051 | 3.669 | 0.462 |
| $\tilde{y}_3$ | 220.010 | -389.100 | 304.510 | -84.608 | 0.798 | 7.513 | -7.051 | 14.979 | -1.382 | 0.273 |
| $\tilde{y}_4$ | -23.244 | 67.989 | -84.608 | 35.151 | -0.740 | 10.029 | 3.669 | -1.382 | 6.594 | 0.223 |
| $\tilde{y}_5$ | 0.119 | -0.349 | 0.798 | -0.740 | 0.200 | 1.263 | 0.462 | 0.273 | 0.223 | 0.188 |

In the second experiment we consider estimation of the parameters of the GB2 distribution under *DGP 1* using two scenarios: observations on $\left(c_i, \tilde{y}_{L,i}, z_i\right)$ and observations on $\left(c_i, \tilde{y}_{L,i}\right)$ only. Experimental results from using the minimum distance estimators (3.8) and (3.17) are given in Table 2. Again, 500 Monte Carlo iterations are used.[15] In addition to the

[15] In all cases, we conducted repeated minimum distance estimation starting with an identity weight matrix and repeating the process 10 times as explained in Section 3. All estimations were done using code written in MATLAB R2013.

GB2 parameters, we also include the Gini coefficient as one of the main quantities of interest in income distribution studies. The left and right panels in the Table display the results for observed and unobserved $\tilde{z}_i$, respectively. In each case, the first column provides the average of the estimated parameters over the Monte-Carlo replications. As it can be seen, they are almost identical to the true values, suggesting unbiasedness. The second, third and fourth columns contain, respectively, (i) the asymptotic variances computed at the true parameter values, (ii) the average of the estimated variances over the replications, and (iii) the Monte Carlo sample variance of the estimated parameters. There are only minor differences between the three sets of variances. The sample variances are slightly larger than their asymptotic counterparts, indicating that there could be some underestimation of the finite sample variation, but, because the differences are not large, overall we can conclude that the estimators perform well in finite samples. We also see that estimation with observed $z_i$'s is more efficient than estimation with unobserved $z_i$'s, but the differences are negligible. At least under the ideal conditions of the Monte Carlo experiment, not knowing the $z_i$'s has little impact.

Table-2: Monte Carlo Experiment Results for *DGP 1*

| | | With data on z | | | | With unknown z | | | |
|---|---|---|---|---|---|---|---|---|---|
| | True Par | Average Est-Par | True Var | Average Est-Var | Variance Est-Par | Average Est-Par | True Var | Average Est-Var | Variance Est-Par |
| $b$ | 100.00 | 100.70 | 32.84 | 34.68 | 36.20 | 100.56 | 34.06 | 36.50 | 37.52 |
| $p$ | 1.0000 | 1.0177 | 0.0122 | 0.0123 | 0.0133 | 1.0155 | 0.0125 | 0.0129 | 0.0136 |
| $q$ | 1.5000 | 1.5338 | 0.0417 | 0.0416 | 0.0464 | 1.5290 | 0.0431 | 0.0440 | 0.0479 |
| $a$ | 1.5000 | 1.4941 | 0.0142 | 0.0133 | 0.0145 | 1.4972 | 0.0145 | 0.0140 | 0.0151 |
| $Gini$ | 0.5326 | 0.5326 | 0.0071 | 0.0067 | 0.0072 | 0.5328 | 0.0073 | 0.0071 | 0.0074 |

Our third experiment is again concerned with estimation of the GB2 parameters and the related Gini coefficient, but this time under *DGP 2*, using the GMM estimator in (4.5), with and without the $z_i$ treated as unknown parameters. The results appear in Table 3 which follows the same format as Table 2 except that, for the unknown $z_i$ case, we report a selection of estimates of the $z_i$. Our conclusions are similar to those from *DGP 1*. There is no evidence that the estimators are biased. The three variances are very similar, with the asymptotic variance and the average of its estimates slightly understating the Monte-Carlo estimated variance. Having to estimate the group boundaries reduces efficiency, but not by much. Comparing the results from *DGP* 1 and *DGP 2*, we see that, despite substantial differences between the models and between the variances of the Lorenz ordinates, the "true" variances for the parameters are identical as predicted by Proposition 7.

Table-3: Monte Carlo Experiment Results for *DGP 2*

| | | Known z | | | | Unknown z | | | |
|---|---|---|---|---|---|---|---|---|---|
| | True Par | Average Est-Par | True Var | Average Est-Var | Variance Est-Par | Average Est-Par | True Var | Average Est-Var | Variance Est-Par |
| $z_1$ | 10.656 | | | | | 10.654 | 0.0159 | 0.0151 | 0.0159 |
| $z_2$ | 17.430 | | | | | 17.438 | 0.0118 | 0.0121 | 0.0118 |
| $z_{10}$ | 70.139 | | | | | 70.153 | 0.0205 | 0.0208 | 0.0205 |
| $z_{18}$ | 236.70 | | | | | 236.69 | 1.36 | 1.63 | 1.36 |
| $z_{19}$ | 343.56 | | | | | 343.65 | 11.39 | 10.47 | 11.50 |
| $b$ | 100.00 | 100.57 | 32.84 | 34.78 | 35.88 | 100.55 | 34.06 | 36.05 | 37.33 |
| $p$ | 1.0000 | 1.0144 | 0.0122 | 0.0123 | 0.0132 | 1.0140 | 0.0125 | 0.0128 | 0.0136 |
| $q$ | 1.5000 | 1.5276 | 0.0417 | 0.0418 | 0.0460 | 1.5268 | 0.0431 | 0.0434 | 0.0477 |
| $a$ | 1.5000 | 1.4976 | 0.0142 | 0.0136 | 0.0146 | 1.4985 | 0.0145 | 0.0140 | 0.0151 |
| $Gini$ | 0.5326 | 0.5328 | 0.0071 | 0.0068 | 0.0072 | 0.5328 | 0.0073 | 0.0071 | 0.0075 |

Our final experiment is designed to compare the performance of the minimum distance estimator in (3.17) with the least squares estimator in (3.20). Using data generated from *DGP 1*, and with unknown group bounds, two scenarios are considered. In the first scenario the data are generated from the GB2 distribution as in the previous experiments, and the parameters of the distribution are estimated using the Lorenz curve specification in (2.8). In the second scenario the data are generated consistently with the SCS Lorenz curve specification in (2.5), and the minimum distance and least squares estimators are used to estimate its parameters.[16] As with the GB2 distribution, 10,000 observations were generated and grouped into 20 groups of equal size. The SCS model parameters $\mu = 40$, $\beta_1 = -2$, $\beta_2 = 0.8$, and $\beta_3 = 3.4$ were chosen as close to the estimates from a real data example in Section 7. They preserve the required properties of a Lorenz curve despite the fact that $\beta_1 = -2$ violates the sufficient condition that all parameters are positive. The support of the distribution is $[0, \infty)$ and the second order moment becomes infinite if $\beta_2 < 0.5$. The results are presented in Table 4 in the same format as those from the earlier experiments. For the least squares variances, we use the "correct" formula in (3.23). The minimum distance estimator has smaller variances in both cases, with the difference being particularly marked for the GB2 distribution. Also, for least squares applied to the GB2 distribution we see some relatively large differences between the true variances, the average of the estimated variances, and the variance of the estimated parameters. These differences can be attributed to the heavy tail of the assumed distribution since some of the least squares estimates approach values that yield an infinite second moment.[17]

There are other questions of interest that can be investigated with Monte Carlo experiments, such as the effect of the number of groups and the effect of the number of observations. We address the effect of the number of groups in Section 8. A Monte Carlo experiment was conducted to examine the effect of increasing the number of observations, yielding results that were consistent with asymptotic theory. The 10000 observations used for the results that we have reported is less than the number from most household surveys.

---

[16] To generate the data according to this Lorenz curve, we use the inverse-cdf method and equation (3.14).

[17] In fact we have discarded 15 of the Monte Carlo draws in reporting the results for LS-GB2 because of their unreasonably large variances.

Table-4: Monte Carlo Results, Minimum Distance vs Least Squares

| | | MD-GB2 | | | | LS-GB2 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | True Par | Average Est-Par | True Var | Average Est-Var | Variance Est-Par | Average Est-Par | True Var | Average Est-Var | Variance Est-Par |
| $b$ | 100.00 | 100.56 | 34.06 | 36.50 | 37.52 | 101.37 | 39.25 | 31.14 | 46.85 |
| $p$ | 1.00 | 1.0155 | 0.0125 | 0.0129 | 0.0136 | 1.0610 | 0.0438 | 0.0353 | 0.0569 |
| $q$ | 1.50 | 1.5290 | 0.0431 | 0.0440 | 0.0479 | 1.5984 | 0.1128 | 0.0756 | 0.1422 |
| $a$ | 1.50 | 1.4972 | 0.0145 | 0.0140 | 0.0151 | 1.4678 | 0.0432 | 0.0254 | 0.0521 |
| $Gini$ | 0.5326 | 0.5328 | 0.0073 | 0.0071 | 0.0074 | 0.5326 | 0.0088 | 0.0070 | 0.0074 |
| | | MD-SCS | | | | LS-SCS | | | |
| | True Par | Average Est-Par | True Var | Average Est-Var | Variance Est-Par | Average Est-Par | True Var | Average Est-Var | Variance Est-Par |
| $\mu$ | 40.00 | 39.98 | 0.14 | 0.15 | 0.14 | 39.98 | 0.14 | 0.15 | 0.14 |
| $\beta_1$ | -2.00 | -2.0924 | 0.2581 | 0.2977 | 0.3065 | -2.0922 | 0.3062 | 0.3384 | 0.3677 |
| $\beta_2$ | 0.800 | 0.8014 | 0.0006 | 0.0006 | 0.0006 | 0.8010 | 0.0007 | 0.0007 | 0.0007 |
| $\beta_3$ | 3.40 | 3.4922 | 0.2533 | 0.2917 | 0.3014 | 3.4921 | 0.2983 | 0.3297 | 0.3594 |
| $Gini$ | 0.4201 | 0.4199 | 0.0035 | 0.0036 | 0.0035 | 0.4199 | 0.0035 | 0.0036 | 0.0035 |

## 7. World Bank Estimation of Lorenz Curves

The World Bank collects and publishes grouped income or expenditure data for a large number of countries over time. These data are usually in the form of population and income shares for a number of groups, typically between 10 and 20. The World Bank website *PovcalNet* uses these data to estimates Lorenz curves and, based on these estimates, provides a variety of poverty and inequality measures.[18] The Lorenz specifications that it uses are the *general quadratic* (GQ) proposed by Villasenor and Arnold (1989) and the *beta* proposed by Kakwani (1980).[19] On *PovcalNet* both of these curves are estimated and the one with the better fit is chosen to compute poverty and inequality measures.[20] In what follows, we first review the GQ and the *beta* Lorenz curves and then estimate these models for several

[18] Recently, the World Bank has started to publish data in the form of 100 groups and to base inequality and poverty estimates on these groups. Earlier groups and their Lorenz curve estimations are still present on the PovcalNet website, however.

[19] The literature abounds with proposals for parametric Lorenz curves. Examples are Kakwani and Podder (1973,1976), Rasche et al. (1980), Gupta (1984), Arnold (1987), Basmann et al. (1990), Chotikapanich (1993), Ryu and Slottje (1996), Sarabia et al. (1999) and Wang et al (2011). The literature is surveyed by Sarabia (2008).

[20] See Datt(1998) for details of the World Bank approach.

countries, comparing the results from the method employed by *PovcalNet* with those from our optimal estimation methods. The results are also compared with those from our preferred Lorenz curve specifications: GB2 and SCS.

The general quadratic Lorenz curve depends on three parameters $\beta_1, \beta_2, \beta_3$ and can be written as

$$\ell_i(1-\ell_i) = \beta_1(c_i^2 - \ell_i) + \beta_2 \ell_i (c_i - 1) + \beta_3(c_i - \ell_i) \tag{7.1}$$

The parameters are often estimated by replacing $\ell_i$ with the observed cumulative income proportions and applying least squares to equation (7.1). This practice is not innocuous, however, since the $\ell_i$ s are present on both sides of the equation. To put (7.1) in the generalised Lorenz framework that we have been using, we let $\mu$ denote the mean of the distribution and define the following

$$e = -\left(1 + \beta_1 + \beta_2 + \beta_3\right), \quad m = \beta_2^2 - 4\beta_1, \quad n = 2\beta_2 e - 4\beta_3 \tag{7.2}$$

The GQ generalised Lorenz curve can then be written as

$$\tilde{y}_i = L(c_i) = -\frac{\mu}{2}\left(\beta_2 c_i + e + \sqrt{m c_i^2 + n c_i + e^2}\right) \tag{7.3}$$

This can be estimated by applying nonlinear least squares which is not prone to the above endogeneity problem, but it is still not optimal.[21] The optimal minimum distance estimator can be applied after using Lemma 1 to find the elements required for the weight matrix, namely, $z_i = F^{-1}(c_i)$ and $m_i^{(2)}$.

One problem with beginning a study with specification of a parametric Lorenz curve rather than a parametric income distribution is that the income distribution implied by the Lorenz curve may only be valid for a limited range of incomes. Having a finite support for income can be a serious drawback with potentially important implications since, in poverty

[21] Because $\beta_2$ appears in (7.3), this equation is not a complete reparameterisation of (7.1). When estimating this Lorenz curve, we minimize the relevant objective function with respect to $\beta_1, \beta_2$ and $\beta_3$; the expressions in (7.2) are convenient for intermediate calculations.

and inequality analysis, the tails of the distribution play important roles. Not all Lorenz curves suffer from this problem; for example, the SCS curve does not, but it is an issue with the GQ. To see the problem, note that using Lemma 1, we can show that

$$F^{-1}(c) = -\frac{\mu}{2}\left[\beta_2 + \frac{2mc+n}{2\sqrt{mc^2+nc+e^2}}\right] \tag{7.4}$$

Inverting the above function and setting $r = \sqrt{n^2 - 4me^2}$ , we obtain the cdf as[22]

$$F(y) = \begin{cases} -\frac{1}{2m}\left[n + \frac{r(\beta_2 + 2y/\mu)}{\sqrt{(\beta_2+2y/\mu)^2 - m}}\right] & \text{if } -\frac{\mu}{2}\left[\frac{n}{2\sqrt{e^2}} + \beta_2\right] \le y \le \frac{\mu}{2}\left[-\frac{(2m+n)}{2\sqrt{m+n+e^2}} - \beta_2\right] \\ 0 \quad \text{otherwise} \end{cases} \tag{7.5}$$

Note that this distribution has a finite support; the upper bound becomes infinite only if $m+n+e^2=0$. It can also be shown that if $m+n+e^2=0$, its second order moment becomes infinite. As we discover, this situation arises with moderately heavy-tailed real data examples.

The second model used by the *PovcalNet* is the beta Lorenz curve

$$\ell_i = c_i - \beta_1 c_i^{\beta_2}(1-c_i)^{\beta_3}, \qquad \text{with} \quad 0 < \beta_1, \beta_2, \beta_3 < 1 \tag{7.6}$$

Under the above conditions on the $\beta_i$s, the corresponding distribution has an infinite support, but in applications its upper tail exhibits inflexibility in the sense that the second order moment often becomes infinite. This can occur if $\beta_2 < 0.5$ or $\beta_3 < 0.5$. In real data examples we frequently obtain estimates for $\beta_3$ that are less than 0.5 even for countries with Gini coefficients as low as 0.3.

In Table 5 we compare estimates obtained using the World Bank's least squares methods with estimates obtained from our proposed minimum distance and GMM methods, for 5 selected countries. Both the GQ and the beta Lorenz curves are considered. In addition

[22] The pdf can be obtained by taking the derivative of (7.5) with respect to *y*.

to the parameter estimates and their standard errors, estimates for the Gini coefficient and the headcount ratio $H$ (using a poverty line of $38/month) are reported. We also include estimates obtained using the GB2 and SCS specifications. Useful diagnostic information that is provided is the supports of the distributions $[y_{\min}, y_{\max}]$, the *J*-statistics for testing the validity of the estimating equations, and the percentage errors in predictions of the incomes of the first and last groups, $S_1$ and $S_{20}$. The data are from 2004 or 2005, extracted from the *PovcalNet* website where we obtained population and income shares for 20 groups and mean income for each country. In all cases the $z_i$ s are not observed, *DGP 1* is assumed, and therefore only the second set of equations in (3.10) is considered. The following abbreviations for the estimators are used in the Table:

(i) OLS-GQ: Least squares estimation applied to (7.1) – the World Bank approach.

(ii) MD-GQ: Minimum distance estimation applied to (7.3).

(iii) NLS-Beta: Nonlinear least squares applied to the generalised Lorenz version of (7.6).

(iv) MD-Beta: Minimum distance estimation applied to the Lorenz curve in (iii).

(v) MD-GB2: Minimum distance estimation applied to the Lorenz curve in (2.9).

(vi) MD-SCS: Minimum distance estimation applied to the Lorenz curve in (2.6).

Table 5: A comparison of Lorenz curve estimation methods for selected countries using 2004 or 2005 data.

## China Urban

| | OLS-GQ | | MD-GQ | | NLS-Beta | MD-Beta | | MD-GB2 | | MD-SCS | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Par** | **SE** | **Par** | **SE** | **Par** | **Par** | **SE** | **Par** | **SE** | **Par** | **SE** |
| $\beta_1/b$ | 0.889 | 0.003 | 0.912 | 0.009 | 0.638 | 0.641 | 0.005 | 108.14 | 4.857 | -0.320 | 0.192 |
| $\beta_2/p$ | -1.260 | 0.010 | -1.344 | 0.018 | 0.935 | 0.935 | 0.003 | 2.369 | 0.391 | 0.708 | 0.027 |
| $\beta_3/q$ | 0.187 | 0.006 | 0.140 | 0.005 | 0.522 | 0.515 | 0.011 | 1.778 | 0.252 | 1.621 | 0.189 |
| $\mu/a$ | NA | NA | 160.95 | 1.216 | 161.87 | 162.53 | 1.426 | 1.842 | 0.168 | 161.47 | 1.316 |
| $Gini$ | 0.347 | 0.000 | 0.349 | 0.004 | 0.353 | 0.352 | 0.004 | 0.345 | 0.004 | 0.347 | 0.004 |
| $H$ | **0.004** | NA | **0.021** | 0.001 | **0.018** | **0.017** | 0.001 | **0.019** | 0.001 | **0.021** | 0.001 |
| $y_{\min}$ | 36.86 | | 31.80 | | 0 | 0 | | 0 | | 0 | |
| $y_{\max}$ | 1487.7 | | 1935.4 | | $\infty$ | $\infty$ | | $\infty$ | | $\infty$ | |
| $J$-Stat | **278.58** | | **85.124** | | **28.79** | **28.40** | | **31.079** | | **53.69** | |
| $S_1$ | **-6.609** | | **2.152** | | **0.631** | **-0.214** | | **0.497** | | **-3.039** | |
| $S_{20}$ | **-0.005** | | **-0.546** | | **0.024** | **0.380** | | **-0.418** | | **-0.222** | |

## Nigeria

| | OLS-GQ | | MD-GQ | | NLS-Beta | MD-Beta | | MD-GB2 | | MD-SCS | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Par** | **SE** | **Par** | **SE** | **Par** | **Par** | **SE** | **Par** | **SE** | **Par** | **SE** |
| $\beta_1/b$ | 0.979 | 0.003 | 0.999 | 0.010 | 0.775 | 0.761 | 0.005 | 32.82 | 3.276 | -2.001 | 0.531 |
| $\beta_2/p$ | -0.932 | 0.015 | -1.068 | 0.026 | 0.974 | 0.960 | 0.002 | 4.88 | 1.508 | 0.812 | 0.025 |
| $\beta_3/q$ | 0.189 | 0.006 | 0.130 | 0.005 | 0.538 | 0.507 | 0.011 | 5.202 | 1.616 | 3.384 | 0.527 |
| $\mu/a$ | NA | NA | 39.764 | 0.329 | 40.272 | 40.946 | 0.436 | 0.893 | 0.150 | 40.003 | 0.353 |
| $Gini$ | 0.400 | 0.000 | 0.401 | 0.003 | 0.404 | 0.414 | 0.005 | 0.401 | 0.004 | 0.401 | 0.003 |
| $H$ | 0.619 | NA | 0.618 | 0.004 | 0.614 | 0.615 | 0.004 | 0.620 | 0.004 | 0.626 | 0.004 |
| $y_{\min}$ | 6.088 | | 4.884 | | 0 | 0 | | 0 | | 0 | |
| $y_{\max}$ | 287.37 | | 326.87 | | $\infty$ | | | $\infty$ | | $\infty$ | |
| $J$-Stat | **399.54** | | **95.164** | | **47.70** | NA | | **39.054** | | **66.779** | |
| $S_1$ | **-12.108** | | **2.439** | | **14.408** | **0.1015** | | **0.099** | | **-4.117** | |
| $S_{20}$ | **-0.006** | | **-0.640** | | **0.077** | **2.818** | | **-0.090** | | **-0.0436** | |

## Pakistan

| | OLS-GQ | | MD-GQ | | NLS-Beta | MD-Beta | | MD-GB2 | | MD-SCS | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Par** | **SE** | **Par** | **SE** | **Par** | **Par** | **SE** | **Par** | **SE** | **Par** | **SE** |
| $\beta_1/b$ | 0.752 | 0.010 | 0.853 | 0.009 | 0.549 | 0.538 | 0.005 | 39.087 | 1.140 | 0.002 | 0.123 |
| $\beta_2/p$ | -1.267 | 0.029 | -1.458 | 0.015 | 0.941 | 0.918 | 0.003 | 1.565 | 0.206 | 0.641 | 0.029 |
| $\beta_3/q$ | 0.269 | 0.018 | 0.167 | 0.005 | 0.462 | 0.399 | 0.014 | 0.691 | 0.067 | 1.186 | 0.121 |
| $\mu/a$ | NA | NA | 63.760 | 0.465 | 66.321 | 68.493 | 0.818 | 3.720 | 0.276 | 65.734 | 0.555 |
| $Gini$ | 0.312 | 0.001 | 0.304 | 0.004 | 0.317 | 0.339 | 0.007 | 0.315 | 0.004 | 0.311 | 0.005 |
| $H$ | 0.232 | NA | 0.235 | 0.004 | 0.229 | 0.219 | 0.004 | 0.221 | 0.004 | 0.216 | 0.004 |
| $y_{\min}$ | 22.77 | | 18.932 | | 0 | 0 | | 0 | | 0 | |
| $y_{\max}$ | 1569.1 | | 1374.2 | | $\infty$ | | | $\infty$ | | $\infty$ | |
| $J$-Stat | **4067.9** | | **370.41** | | NA | NA | | **20.221** | | **15.58** | |
| $S_1$ | **-15.207** | | **3.101** | | **8.414** | **-0.309** | | **-0.200** | | **0.2099** | |
| $S_{20}$ | **0.004** | | **-3.041** | | **0.061** | **4.236** | | **0.511** | | **-0.039** | |

| **Kenya** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **OLS-GQ** | | **MD-GQ** | | **NLS-Beta** | **MD-Beta** | | **MD-GB2** | | **MD-SCS** | |
| | **Par** | **SE** | **Par** | **SE** | **Par** | **Par** | **SE** | **Par** | **SE** | **Par** | **SE** |
| $\beta_1/b$ | 0.755 | 0.008 | Infinite second order moment No estimates are reported | | 0.800 | 0.779 | 0.005 | 35.521 | 1.239 | -0.038 | 0.178 |
| $\beta_2/p$ | -0.485 | 0.048 | | | 0.975 | 0.953 | 0.002 | 1.337 | 0.180 | 0.544 | 0.042 |
| $\beta_3/q$ | 0.222 | 0.016 | | | 0.392 | 0.322 | 0.015 | 1.015 | 0.133 | 1.433 | 0.173 |
| $\mu/a$ | NA | NA | | | 66.949 | 72.019 | 1.773 | 1.967 | 0.172 | 65.592 | 1.287 |
| $Gini$ | 0.456 | 0.010 | | | 0.488 | 0.523 | 0.011 | 0.482 | 0.009 | 0.477 | 0.009 |
| $H$ | 0.438 | NA | | | 0.441 | 0.438 | 0.004 | 0.440 | 0.004 | 0.439 | 0.004 |
| $y_{\min}$ | 29.188 | | | | 0 | 0 | | 0 | | 0 | |
| $y_{\max}$ | 10615 | | | | $\infty$ | | | $\infty$ | | $\infty$ | |
| $J$-Stat | **NA** | | | | NA | NA | | **86.582** | | **81.287** | |
| $S_1$ | **39.686** | | | | **31.272** | **-0.248** | | **0.985** | | **1.969** | |
| $S_{20}$ | **-0.002** | | | | **0.055** | **8.701** | | **1.091** | | **0.186** | |

| **Iran** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **OLS-GQ** | | **MD-GQ** | | **NLS-Beta** | **MD-Beta** | | **MD-GB2** | | **MS-SCS** | |
| | **Par** | **SE** | **Par** | **SE** | **Par** | **Par** | **SE** | **Par** | **SE** | **Par** | **SE** |
| $\beta_1/b$ | 0.888 | 0.004 | 0.922 | 0.010 | 0.709 | 0.698 | 0.005 | 118.99 | 7.847 | -0.822 | 0.263 |
| $\beta_2/p$ | -1.030 | 0.018 | -1.190 | 0.022 | 0.957 | 0.946 | 0.003 | 3.059 | 0.602 | 0.734 | 0.026 |
| $\beta_3/q$ | 0.212 | 0.008 | 0.135 | 0.005 | 0.504 | 0.484 | 0.012 | 2.258 | 0.377 | 2.147 | 0.260 |
| $\mu/a$ | NA | NA | 195.99 | 1.682 | 199.09 | 200.93 | 2.143 | 1.426 | 0.149 | 197.20 | 1.781 |
| $Gini$ | 0.383 | 0.000 | 0.385 | 0.004 | 0.388 | 0.393 | 0.005 | 0.383 | 0.004 | 0.383 | 0.004 |
| $H$ | **0.000** | NA | **0.023** | 0.001 | **0.012** | **0.016** | 0.001 | **0.020** | 0.001 | **0.020** | 0.001 |
| $y_{\min}$ | 38.822 | | 30.463 | | 0 | 0 | | 0 | | 0 | |
| $y_{\max}$ | 1877.9 | | 2718.8 | | $\infty$ | | | $\infty$ | | $\infty$ | |
| $J$-Stat | **820** | | **187.97** | | NA | NA | | **33.744** | | **56.479** | |
| $S_1$ | **-12.613** | | **3.846** | | **8.505** | **0.0867** | | **0.497** | | **-3.196** | |
| $S_{20}$ | **-0.006** | | **-0.851** | | **0.032** | **1.634** | | **-0.357** | | **-0.239** | |

Notes: OLS-GQ parameters and standard errors are obtained by application of OLS to (7.1). In this case we do not estimate $\mu$. No estimates are reported for MD-GQ for Kenya because its second moment is infinite. MD-GQ, MD-GB2 and MD-SCS estimates are obtained by applying the minimum distance method (3.17). Their standard errors use (3.18). Since sample sizes are not provided on *PovcalNet*, we assume $T = 10000$. NLS-Beta is obtained by application of nonlinear least squares to the beta generalised Lorenz curve, but no NLS standard errors are reported since such standard errors do not have a sound statistical basis. MD-Beta is obtained by application of the minimum distance method, but, since the second order moment often turns out to be infinite, we used $c_N$ = 0.99999 (instead of 1) to make estimation feasible, and to enable us to report some numbers. This alternative is not ideal, but an alternative inference method is not available. A *J*-statistic $J = T\left(\tilde{\mathbf{y}}_L - \mathbf{L}(\mathbf{c};\hat{\boldsymbol{\phi}})\right)' \hat{\boldsymbol{\Omega}}_{L,22}^{-1}\left(\tilde{\mathbf{y}}_L - \mathbf{L}(\mathbf{c};\hat{\boldsymbol{\phi}})\right)$ is computed for the minimum distance cases (except for the Beta Lorenz curve where it suffered from heavy-tail issues) and also for OLS-GQ and NLS-Beta. In these latter 2 cases we used the estimated parameters to compute a weight matrix as described by MD theory. The critical value for the *J*-statistic at a 0.05 level of significance is $\chi^2_{16} = 26.3$. $S_1$ and $S_{20}$ are the percentage errors in prediction of the first and last groups' income.

The results can be summarised as follows:

1. Estimates for $\beta_1, \beta_2, \beta_3$ from least squares and from minimum distance estimation are often sufficiently different to have important implications, especially for the GQ model. The least squares estimates for the GQ and beta Lorenz curves lead to relatively large percentage errors for the estimated incomes in the first group, $S_1$. This is particularly pronounced for the GQ curve which has the added problem of a finite support. It can lead to poor estimates for the head-count ratio and most likely even poorer estimates for other poverty measures.[23] China-Urban and Iran are good examples.[24] In these examples the minimum distance method improves the estimated support and provides much more reasonable values for the head-count ratio. For China-Urban, OLS-GQ produces a head count ratio of 0.004, $y_{\min} = 36.86$, $y_{\max} = 1487.7$ and $S_1 = -6.61$, while MD-GQ leads to $H = 0.021$, $y_{\min} = 31.80$, $y_{\max} = 1935.4$ and $S_1 = 2.15$. For Iran, the corresponding numbers are $H = 0$, $x_{\min} = 38.86$, $x_{\max} = 1877.9$ and $S_1 = -12.6$ for OLS-GQ and $H = 0.023$, $y_{\min} = 30.463$, $y_{\max} = 2718.8$ and $S_1 = 3.84$ for MD-GQ. In both cases, the estimated supports of the distributions from minimum distance methods are more reasonable and for head-count ratios not substantially different from GB2 or SCS.
2. The standard errors from the two methods of estimation are very different, but cannot really be compared since the least squares ones do not use (3.23).
3. The GQ and especially the beta Lorenz curve can suffer from heavy-tail "inflexibility". For example, with GQ the second order moment becomes infinite for the case of Kenya with a Gini coefficient of 0.47 and thus no estimates are reported for MD-GQ. With the beta Lorenz curve we have infinite second order moments for all cases except China Urban. The reported results are from estimating the model using $c_N = 0.99999$ instead

[23] This has been realized in *PovcalNet* and estimates from a beta Lorenz curve are used in such cases.
[24] This problem also existed for a number of other countries. We have limited our reported results to 5 countries.

of 1 to ensure the existence of a finite second moment.[25] There were no problems with the GB2 and SCS curves for the examples considered here, but for countries with higher degrees of inequality such as South Africa or Brazil even these functions suffer from the same issue.[26]

4. Unlike the poverty measures, estimates of the Gini coefficient are not overly sensitive to the method of estimation or the model used.

5. An overall measure of the goodness of fit of the models is the *J*-statistic given by $J = T\left(\tilde{\mathbf{y}}_L - \mathbf{L}(\mathbf{c};\hat{\boldsymbol{\phi}})\right)' \hat{\boldsymbol{\Omega}}_{L,22}^{-1}\left(\tilde{\mathbf{y}}_L - \mathbf{L}(\mathbf{c};\hat{\boldsymbol{\phi}})\right)$. If the moment conditions are valid, $J \sim \chi^2_{(N-d)}$ where *d* is the number of estimated parameters and *N* is the number of moment conditions. Smaller values of *J* are an indication of better fitting Lorenz curves. Based on this criterion, GB2 performs better than SCS in three of the cases and is slightly worse in the other two cases. They both have substantially smaller *J*-statistics compared to the GQ model[27].

Based on our theoretical analyses, the Monte Carlo experiments, and real data estimations we make the following recommendations to improve the current practice of Lorenz curve estimation by the World Bank and others.

1. Both the GQ and beta Lorenz curves have undesirable properties. The GQ has an inflexible bounded support resulting in poor estimates for poverty measures and the beta Lorenz curve has an “inflexible” upper tail. Other Lorenz curve candidates without such problems, such as the GB2 or SCS models, are preferable. These two models, although not linear, are still easy to estimate and exhibit good performance.

[25] Any finite positive definite weight matrix (instead of the optimal weight matrix **W)** will give a consistent estimator for the parameters, but inference methods still need to be worked out for when the variance is infinite. Exposing this problem reinforces the need for a sound statistical framework to reveal the difficulty and to provide a foundation for investigating a solution.

[26] In such cases, we may prefer to use a distribution with more flexible tails such as a mixture of lognormals.

[27] *Povcal* decides between GQ and *beta* Lorenz curves based on their in-sample income share prediction performance where GQ often performs better. A more satisfactory criterion is the J-stat which seems to support to the *beta* Lorenz curves over GQ more often.

2. The proposed minimum distance estimation method should be used instead of least square methods since it is more efficient and provides sound standard errors.
3. If the GQ model is being used, say because of its simplicity, the minimum distance estimator should be used instead of OLS. It provides better estimates for the support of the distribution and more accurate poverty measure estimates.
4. It might be worthwhile to search for a Lorenz curve specification that can be made linear in parameters but has still good tail properties.
5. If possible, published data should include the sample size, and the variance and boundaries of the groups. Having this information makes parametric inference easier and inference using distribution-free approaches feasible.

## 8. Estimation of Lorenz Curves with Individual Data

An interesting question that we have not yet addressed is what happens if the number of groups is increased? A complete and formal answer to this question is beyond the scope of this paper. However, in what follows we first provide a Monte Carlo experiment to show the effect of a common type of increase in the number of groups and then we study perhaps the more interesting case where we consider using individual data to estimate a Lorenz curve or a quantile function. We discover that these approaches are viable alternatives to maximum likelihood.

To see how the variances of estimated parameters change as the number of groups increases, a Monte Carlo experiment is performed where the data again consist of 10000 observations, drawn from a GB2 distribution with $b = 100$, $p = 1$, $q = 1.5$ and $a = 1.5$; they are grouped into 10, 20 and 100 groups using *DGP* 1 with the proportion of observations being the same for each group. Table 6 presents the results from estimation of these models where the columns headed "*Par*", "*Var1*" and "*Var2*" contain the averages of the estimated parameters, the averages of the estimated variances, and the variances computed at the true values of the parameters, respectively. Maximum likelihood results based on individual data are also presented.[28] Two conclusions are evident. The first is that the variances decrease as the number of groups increases. The variances for 10 groups are between 48% and 63% higher than those from maximum likelihood; those from 100 groups are from 9% to 13% higher. The second conclusion is that the averages of the estimated variances are always greater than the asymptotic variances evaluated at the true parameters, but the differences between them decreases as the number of groups increases.

Table-6 Monte Carlo Experiment results with different group numbers

| | MD with 10 Groups | | | MD with 20 Groups | | | MD with 100 Groups | | | ML with Individual Data | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Par | Var1 | Var2 | Par | Var1 | Var2 | Par | Var1 | Var2 | Par | Var1 | Var2 |
| ***b*** | 99.775 | 41.770 | 38.113 | 99.748 | 35.803 | 34.907 | 100.025 | 28.478 | 26.627 | 99.871 | 24.943 | 23.875 |
| ***p*** | 1.0150 | 0.0161 | 0.0142 | 1.0129 | 0.0136 | 0.0128 | 1.0151 | 0.0113 | 0.0105 | 1.0122 | 0.0102 | 0.0096 |
| ***q*** | 1.5184 | 0.0553 | 0.0487 | 1.5153 | 0.0467 | 0.0443 | 1.5220 | 0.0368 | 0.0338 | 1.5154 | 0.0320 | 0.0299 |
| ***a*** | 1.5017 | 0.0185 | 0.0163 | 1.5042 | 0.0153 | 0.0149 | 1.4982 | 0.0121 | 0.0120 | 1.4998 | 0.0111 | 0.0110 |

Now, suppose we have individual data on income denoted by $\left(y_1,\ldots,y_T\right)$ and consider three parametric representations of a distribution, namely, its density function $f(y;\phi)$, its quantile function $F^{-1}(c;\phi)$ and the Lorenz curve $L(c;\phi)$. We can estimate $\phi$ using at least three different approaches:

*a)* Density Function Estimation: This is the most common approach where a density function $f(y;\phi)$ is estimated using maximum likelihood.

*b)* Quantile Function Estimation: The sample is arranged in order from the lowest to the highest forming $\tilde{\mathbf{z}}'=\left(y_{(1)},\ldots,y_{(T-1)}\right)$, the cumulative population proportion is

[28] To compute the asymptotic variances for the maximum likelihood estimator at the true parameter values we used the GB2 information matrix derived by Brazauskas (2002).

defined as $\mathbf{c}' = \left(1/T, 2/T, \ldots, (T-1)/T\right)$, and then a quantile function $\tilde{z}_i = F^{-1}(c_i; \boldsymbol{\phi})$ is estimated as we describe below.[29]

*c)* Lorenz Curve Estimation: Income levels are ordered, their cumulative sums $\tilde{y}_i$s and the population proportions are computed, and then a Lorenz curve $\tilde{y}_i = L(c_i; \boldsymbol{\theta})$ is estimated as explained in Section 3 and below.

The first question to ask is why one might be interested in doing (*b*) or (*c*) instead of (*a*) when it is known that maximum likelihood provides the most efficient estimator. The answer is that under some circumstances, cases (*b*) and (*c*) might be easier to perform. For example, it might be possible to specify quantile or Lorenz functions that are linear in the parameters. Also, in some cases it might be more natural to estimate a quantile function or a Lorenz curve. An example of a study along these lines is Ryu and Slottje (1996). They propose two flexible functional forms, one for a quantile function and one for a Lorenz curve. For the quantile function, they propose the power series $F^{-1}(c_i; \boldsymbol{\phi}) = \exp\left\{\phi_0 + \phi_1 c_i + .... + \phi_l c_i^l\right\}$; for a Lorenz curve they use Bernstein polynomials. However, they estimate the models using simple least squares ignoring all the correlations between the error terms. The theory we developed in Section 3 can be used for optimal estimation of the models with individual data and is not substantially more difficult. Consider first the quantile estimation problem.

***Proposition 8:*** Under some regularity conditions and data constructed as in case (b) above, $\hat{\boldsymbol{\phi}}$ defined by

$$\begin{aligned}\hat{\boldsymbol{\phi}} &= \arg\min_{\boldsymbol{\phi}} \left(\tilde{\mathbf{z}} - \mathbf{F}^{-1}(\mathbf{c}; \boldsymbol{\phi})\right)' \tilde{\mathbf{W}}_{L,11} \left(\tilde{\mathbf{z}} - \mathbf{F}^{-1}(\mathbf{c}; \boldsymbol{\phi})\right) \\ &= \arg\min_{\boldsymbol{\phi}} N \sum_{i=0}^{N-1} \left\{ \frac{\tilde{z}_{i+1} - F^{-1}(c_{i+1}; \boldsymbol{\phi})}{G_{i+1}} - \frac{\tilde{z}_i - F^{-1}(c_i; \boldsymbol{\phi})}{G_i} \right\}^2\end{aligned} \tag{8.1}$$

*(i)* is a consistent estimator of $\boldsymbol{\phi}_0$

[29] Any real-valued non-decreasing left continuous function with domain (0, 1) can be a quantile function.

*(ii)* and is asymptotically normal

$$\sqrt{N}\left(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0\right)\xrightarrow{d}N(\mathbf{0},\mathbf{V}) \quad \text{with} \quad \mathbf{V}=\left(\frac{\partial \mathbf{F}'^{-1}}{\partial \boldsymbol{\phi}}\mathbf{W}_{L,11}\frac{\partial \mathbf{F}^{-1}}{\partial \boldsymbol{\phi}'}\right)^{-1} \tag{8.2}$$

where

$$\tilde{\mathbf{W}}_{L,11}=N\begin{bmatrix} \frac{2}{G_1^2} & \frac{-1}{G_1G_2} & & \\ \frac{-1}{G_1G_2} & \ddots & \ddots & \\ & \ddots & \ddots & \frac{-1}{G_{N-2}G_{N-1}} \\ & & \frac{-1}{G_{N-2}G_{N-1}} & \frac{2}{G_{N-1}^2} \end{bmatrix} \tag{8.3}$$

$\left(\tilde{z}_i-F^{-1}(c_i;\boldsymbol{\phi})\right)/G_i=0$ for $i=0,N$ , $G_i=\partial F^{-1}\left(c_i;\tilde{\boldsymbol{\phi}}\right)/\partial c_i$ for $i=1,\ldots,N-1$ , and blank elements in the weight matrix are zeros. Note that with individual data, $N=T$ .

*Proof*: See Appendix 5.

Since $\mathbf{W}_{L,11}$ has a simple closed tri-diagonal form resulting in (8.1), applying optimal minimum distance estimation is straightforward and no more difficult than least squares or maximum likelihood density estimation.

Defining $\mathbf{c}'=\left(1/T,2/T,\ldots,1\right)$ and $\tilde{y}_i$'s as described in case (*c*) above, one can think of using (3.17) to estimate a Lorenz curve $\mathbf{L}(\mathbf{c};\boldsymbol{\phi})$ . That is,

$$\hat{\boldsymbol{\phi}}=\arg\min_{\boldsymbol{\phi}}\left(\tilde{\mathbf{y}}-\mathbf{L}(\mathbf{c};\boldsymbol{\phi})\right)'\tilde{\boldsymbol{\Omega}}_{L,22}^{-1}\left(\tilde{\mathbf{y}}-\mathbf{L}(\mathbf{c};\boldsymbol{\phi})\right) \tag{8.4}$$

One potential concern with this estimator is the large size of $\tilde{\boldsymbol{\Omega}}_{22}$ since, with individual data, $\tilde{\boldsymbol{\Omega}}_{L,22}$ is a $T\times T$ matrix.[30] However, we showed in Proposition 6 that inverting this matrix is

[30] Another concern from a theoretical standpoint is that the number of moments increases as the number of observations increases. This can create technical difficulties when proving consistency and asymptotic normality of the estimator (see e.g., the proof for the quantile estimator). We leave the details for further work, but simulations with several models and for several different values for the parameters indicate that the asymptotic theory developed for a fixed number of groups works for this case as well.

feasible since it can be turned into inversion of a tri-diagonal matrix. Now we present the following proposition that indicates that the asymptotic variances of all three estimators converge to the Cramer-Rao lower bound.

***Proposition 9:*** The asymptotic covariance matrices of the following three estimators for $\phi$ all converge to the Cramer-Rao lower bound.

Maximum likelihood $\qquad \hat{\phi}_{ML} = \arg\max_{\phi} \sum_{i=1}^{N} \log f\left(y_i;\phi\right)$

Quantile function $\qquad \hat{\phi}_{QF} = \arg\min_{\phi} N \sum_{i=0}^{N-1} \left\{ \frac{\tilde{z}_{i+1} - F^{-1}(c_{i+1};\phi)}{G_{i+1}} - \frac{\tilde{z}_i - F^{-1}(c_i;\phi)}{G_i} \right\}^2$

Lorenz curve $\qquad \hat{\phi}_{LC} = \arg\min_{\phi} \left(\tilde{\mathbf{y}} - \mathbf{L}(\mathbf{c};\phi)\right)' \tilde{\Omega}_{L,22}^{-1} \left(\tilde{\mathbf{y}} - \mathbf{L}(\mathbf{c};\phi)\right)$

*Proof*: See Appendix 5.

To illustrate the feasibility of the above estimators with single observations and to study the finite sample relevance of Proposition 9, we applied the three estimators to a data set comprising 24,687 single observations on urban Indonesian monthly household expenditure taken from a 2005 survey.[31] The Dagum distribution, a special case of the GB2 distribution with $q = 1$, was used as the parametric specification.[32] Its density function, quantile function, and Lorenz curve are, respectively,

$$f(y;\phi) = \frac{apy^{ap-1}}{b^{ap}[1 + (y/b)^a]^{p+1}}$$

$$F^{-1}(c_i;\phi) = b\left(c_i^{-1/p} - 1\right)^{-1/a}$$

$$L(c_i;\phi) = bp\mathrm{B}\left(p + 1/a, 1 - 1/a\right) B\left(c_i^{1/p}; p + 1/a, 1 - 1/a\right)$$

[31] The data were kindly provided by Ari Handayani of Monash University. They have been adjusted using household equivalence scales.

[32] We also estimated the more general GB2 specification and the results can be provided upon request but the estimated parameters had very large variances in all cases of density, quantile and Lorenz curve estimation.

The upper and lower left panels of Table 7 provide the parameter estimates and their covariance matrix from maximum likelihood and from the quantile model estimator in (8.1). Estimation of the quantile model was not substantially more involved since its weight matrix is tri-diagonal. Using (8.4) to estimate the Lorenz model by itself is more challenging, however. The weight matrix is not tri-diagonal and estimation involves inversion of a tri-diagonal matrix. Using Proposition 6 and MATLAB sparse matrix options we were able to estimate this model using 10000 observations randomly drawn from the full sample and with non-optimised code.[33] Thus, optimal minimum distance estimation of the Lorenz only model with large data sets, although more challenging, is completely feasible. The results appear in the top right panel of Table 7. Paradoxically, estimation is easier when the Lorenz and quantile functions are combined as in (3.8), because, in this case, we have a closed form block-tri-diagonal weight matrix from Proposition 2. All the observations were included for this estimation; the results appear in the lower right panel of Table 7.

Table 7: Quantile, Density and Lorenz Curve Estimation Results from Single Observations using Indonesian Data and the Dagum Distribution

| | **Density** | | | | **Lorenz with 10000 Observation** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Par** | **Covariance Matrix** | | | **Par** | **Covariance Matrix** | | |
| ***b*** | 203.350 | 26.367 | -0.541 | 0.0818 | 207.700 | 64.303 | -1.241 | 0.202 |
| ***p*** | 2.699 | -0.541 | 0.0114 | -0.00162 | 2.595 | -1.241 | 0.0246 | -0.00381 |
| ***a*** | 2.224 | 0.0818 | -0.00162 | 0.00037 | 2.232 | 0.202 | -0.00381 | 0.00091 |
| | **Quantile** | | | | **Joint Quantile & Lorenz** | | | |
| | **Par** | **Covariance Matrix** | | | **Par** | **Covariance Matrix** | | |
| ***b*** | 202.650 | 26.438 | -0.547 | 0.0818 | 202.670 | 26.338 | -0.549 | 0.0775 |
| ***p*** | 2.725 | -0.547 | 0.0116 | -0.00167 | 2.714 | -0.549 | 0.0117 | -0.00158 |
| ***a*** | 2.218 | 0.0818 | -0.00167 | 0.00036 | 2.221 | 0.0775 | -0.00158 | 0.00033 |

[33] The estimation took less than 3 minutes on a 4 year old HP EliteBook 2540p Laptop with 4 Gigabytes of RAM and an Intel Core i7-640LM 2.13GHz Dual-Core Processor. Given our computer code, for estimation of the Lorenz only model with more than 10000 observations we needed higher RAM to store the weight matrix elements.

The results in Table 7 are in line with Proposition 9. That is, the parameters and the estimated variances are very close to those of the benchmark maximum likelihood. Note that, as expected, when only 10,000 observations are used the estimated parameters differ more and the variances are substantially higher. If these variances are adjusted by the number of observations, they become similar to those from the other cases. Thus, the techniques proposed in the paper are relevant not just for grouped data, but are also useful when the focus is on specification of a quantile function or a Lorenz curve to be estimated using individual data.

## 9. Conclusion

The Lorenz curve has become one of the main tools for poverty and inequality analysis because of its theoretical appeal and because of data availability. However, so far methods for parametric estimation of Lorenz curves have been adhoc. In this paper, we developed statistically sound approaches to Lorenz curve estimation and inference. We performed simulation experiments and real data examples to show that the proposed methods work well and can improve on conventional approaches. In particular we showed that World Bank's estimates for poverty and inequality measures can be improved by using the proposed methods. Based on the experiments and the theoretical results some recommendations were made to improve current practices by the World Bank and other institutions. There is scope for extending the current work by incorporating features such as censoring, trimming, heavy tails, measurement errors, and survey sampling designs that are recognised to have important implications for welfare measurement.

## References


Arnold, B. (1987). *Majorization and the Lorenz Order*, Lecture Notes in Statistics 43, Berlin and New York: Springer.

Atkinson, A. (1970). On the Measurement of Inequality, *Journal of Economic Theory*, 2, 244–263.

Basmann, R., Hayes, K., Slottje, D. and Johnson J. (1990) A General Functional Form for Approximating the Lorenz Curve, *Journal of Econometrics*, 43, 77-90.

Beach, C. and R. Davidson (1983). Distribution-Free Statistical Inference with Lorenz Curves and Income Shares. *Review of Economic Studies* 50, 723-725.

Bishop, J., S. Chakraborti, and P. Thistle (1989). Asymptotically Distribution-Free Statistical Inference for Generalized Lorenz Curves, *Review of Economics and Statistics*, 71, 725-727.

Brazauskas, V. (2002). Fisher Information Matrix for the Feller–Pareto Distribution, *Statistics and Probability Letters*, 59, 159–167.

Chen, S. and M. Ravallion, (2011). The Developing World is Poorer than we Thought, but no Less Successful in the Fight against Poverty, *Quarterly Journal of Economics*, 125, 1577-1625.

Chotikapanich, D. (1993). A Comparison of Alternative Functional Forms for the Lorenz Curve, *Economics Letters*, 41, 129-138.

Chotikapanich, D. and W. Griffiths (2002). Estimating Lorenz Curves Using a Dirichlet Distribution, *Journal of Business and Economic Statistics*, 20, 290-295.

Chotikapanich, D., W. Griffiths, and D.S.P. Rao (2007). Estimating and Combining National Income Distributions using Limited Data, *Journal of Business and Economic Statistics*, 25, 97-109.

Cowell, F. and M. Victoria-Feser (2002). Welfare Rankings in the Presence of Contaminated Data, *Econometrica, 70*, 1221–1233.

Cowell, F. and M. Victoria-Feser (2007). Robust Stochastic Dominance: A Semi-Parametric Approach, *Journal of Economic Inequality,* 5, 21–37.

Cowell, F. and M. Victoria-Feser (2008). Modelling Lorenz Curves: Robust and Semi-Parametric Issues, in *Modeling Income Distribuions and Lorenz Curves*, ed. by D. Chotikapanich, pp. 241–225. Springer, New-York.

Datt, G. (1998). Computational Tools for Poverty Measurement and Analysis. Washington D.C.: International Food Policy Research Institute (IFPRI) Discussion Paper No.50 (Food Consumption and Nutrition Division).

Gastwirth, J. (1971). A General Definition of the Lorenz Curve. *Econometrica*, 39, 1037–1039.

Gastwirth, J. (1972). The Estimation of the Lorenz Curve and Gini Index. *The Review of Economics and Statistics*, 54, 306–316

Gastwirth, J. and M. Gail. (1985). Simple Asymptotically Distribution-Free Methods Comparing Lorenz Curves and Gini Indices Obtained from Complete Data, in R. Basmann and G. Rhodes, Jr., editors, *Advances in Econometrics*, 4, JAI Press, Greenwich, CT.

Goldie, C. (1977). Convergence Theorems for Empirical Lorenz Curves and Their Inverses, *Advances in Applied Probability*, 9, 765–791.

Griffiths, W. and G. Hajargasht (2015). On GMM Estimation of Distributions from Grouped Data, *Economics Letters*, 126, 122-126.

Gupta, M. (1984). Functional Form for Estimating the Lorenz Curve, *Econometrica*, 52, 1313-1314.

Hajargasht, G., Griffiths, W., Brice, J., Rao, D. S. P. and Chotikapanich, D. (2012). Inference for Income Distributions Using Grouped Data. *Journal of Business of Economic Statistics*, 30, 563-575.

Iritani, J., and Kuga, K. (1983). Duality between the Lorenz Curves and the Income Distribution Functions. *Economic Studies Quarterly*, 34, 9–21.

Kakwani, N. and N. Podder (1973). On the Estimation of Lorenz Curves from Grouped Observations, *International Economic Review*, 14, 278-292.

Kakwani, N. and N. Podder (1976). Efficient Estimation of the Lorenz Curve and Associated Inequality Measures from Grouped Observations, *Econometrica*, 44, 137-148.

Kakwani, N. (1980). On a Class of Poverty Measures, *Econometrica*, 48, 437–446.

Kleiber, C. and S. Kotz, (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*, John Wiley and Sons, New York.

Lorenz, M.O. (1905). Methods of Measuring the Concentration of Wealth, *Quarterly Publications of the American Statistical Association,* 9 (New Series, No. 70), 209–219.

McDonald, J. (1984). Some Generalized Functions of the Size Distribution of Income. *Econometrica* 52, 647-663.

McDonald, J. and M. Ransom (2008). The Generalized Beta Distribution as a Model for the Distribution of Income: Estimation of Related Measures of Inequality. In D. Chotikapanich (ed.), *Modeling Income Distributions and Lorenz Curves*, Springer, New York, 147-166.

Newey W. and D. McFadden (1994), Large Sample Estimation and Hypothesis Testing, *Ch 36 in Handbook of Econometrics*, Amsterdam, Elsevier, Vol 4, 2113-2245.

Ortega, P., A. Fernández, A. Ladoux, and A. García (1991). A New Functional Form for Estimating Lorenz Curves, *Review of Income and Wealth*, 37, 447-452.

Rasche, R., J. Gaffney, A. Koo, and N. Obst (1980). Functional Forms for Estimating the Lorenz Curve, *Econometrica*, 48, 1061-1064.

Ryu, H. and D. Slottje. (1996). Two Flexible Functional Form Approaches for Approximating the Lorenz Curve, *Journal of Econometrics* 72: 251-274.

Ryu, H and D. Slottje (1998). Parametric Approximation of the Lorenz Curve, in *Handbook on Income Inequality Measurement* (Kluwer, Boston), 291-314.

Sarabia, J. (2008). Parametric Lorenz Curves: Models and Applications, in *Modeling Income Distribuions and Lorenz Curves*, ed. by D. Chotikapanich, pp. 167–190. Springer, New-York.

Sarabia, J., E. Castillo and D. Slottje (1999). An Ordered Family of Lorenz Curves, *Journal of Econometrics*, 91, 43-60.

Sarabia, J., F. Prieto and V. Jordá (2015). About the Hyperbolic Lorenz Curve, *Economics Letters*, 136, 42-45.

Shorrocks, A. (1983). Ranking Income Distributions, *Economica*, 50, 3–17.

Thistle, P.D. (1989). Duality between Generalized Lorenz Curves and Distribution Functions, *Economic Studies Quarterly*, 40, 183–187.

Villasenor, J., and B. Arnold (1989). Elliptical Lorenz curves. *Journal of Econometrics,* 40 327–338.

Wang, Z., Y.-K. Ng and R. Smyth (2011). A General Method for Creating Lorenz Curves, *The Review of Income and Wealth*, 57, 561-582.

Wang, Z. and R. Smyth (2015a). A Piecewise Method for Estimating the Lorenz Curve, *Economics Letters*, 129, 45-48.

Wang, Z. and R. Smyth (2015b). A Hybrid Method for Creating Lorenz Curves, *Economics Letters*, 133, 59-63.

Wu X. (2006). Inference and Density Estimation with Interval Statistics, www.agecon2.tamu.edu/people/faculty/wu-ximing/agecon2/.../interval.pdf

**Appendix 1**

**i. Derivation of** $\boldsymbol{\Omega}_{L,12} = \text{cov}\left(\sqrt{T}\tilde{\mathbf{z}}, \sqrt{T}\tilde{\mathbf{y}}_L\right)$

To our knowledge $\text{cov}\left(\sqrt{T}\tilde{\mathbf{z}}, \sqrt{T}\tilde{\mathbf{y}}_L\right)$ has not been provided in the literature although its derivation is straightforward and similar to those for $\text{var}\left(\sqrt{T}\tilde{\mathbf{z}}\right)$ and $\text{var}\left(\sqrt{T}\tilde{\mathbf{y}}_L\right)$. We follow Cowell and Victoria-Feser (2007) and use influence functions for the derivation. Denoting the influence function corresponding to $\tilde{z}_i$ by $IF(y;z_i)$ and that corresponding to $\tilde{y}_{L,j}$ by $IF(y;m_i)$, we have

$$\text{cov}\left(\sqrt{T}\tilde{z}_i, \sqrt{T}\tilde{y}_{L,j}\right) = \int IF(y;z_i) IF(y;m_i) f(y) dy \tag{A1.1}$$

It has also been shown that (see Cowell and Victoria-Feser 2007)

$$IF(y;z_i) = \frac{c_i - g(y \le z_i)}{f(z_i)} \tag{A1.2}$$

$$IF(y;m_i) = c_i z_i - m_i + g(y \le z_i)(y - z_i) \tag{A.1.3}$$

where $m_i = L(c_i)$ and $g(y \le z_i) = 1$ when $y \le z_i$, and is 0 otherwise. Using the above facts, we have, for $i = 1, \ldots, N-1$ and $j = 1, \ldots, N$,

(a) if $i \le j$

$$\begin{aligned} \text{cov}\left(\sqrt{T}\tilde{z}_i, \sqrt{T}\tilde{y}_{L,j}\right) &= \int \frac{c_i - g(y \le z_i)}{f(z_i)} \left(c_j z_j - m_j + g(y \le z_j)(y - z_j)\right) f(y) dy \\ &= \frac{c_i(m_j - z_j c_j) - (m_i - z_j c_i)}{f(z_i)} \end{aligned} \tag{A1.4}$$

(b) if $i \ge j$

$$\begin{aligned} \text{cov}\left(\sqrt{T}\tilde{z}_i, \sqrt{T}\tilde{y}_{L,j}\right) &= \int \frac{c_i - g(y \le z_i)}{f(z_i)} \left(c_j z_j - m_j + g(y \le z_j)(y - z_j)\right) f(y) dy \\ &= \frac{(c_i - 1)(m_j - z_j c_j)}{f(z_i)} \end{aligned} \tag{A1.5}$$

## ii. Regularity conditions and proof of Proposition 1

Suppose we have a random sample of size $T$ from a population with generalised Lorenz curve $L(c,\phi_0)$ and that the following assumptions are valid:

1) $L(c,\phi_0)$ satisfies the conditions of a generalised Lorenz curve, is three times differentiable with respect to $c$, $\partial^2 L(c;\phi_0)\big/\partial c^2 > 0$ and $m^{(2)} = \int_0^1 \left(\partial L(x;\phi_0)\big/\partial x\right)^2 dx$ is finite.

2) Compactness: $\phi_0 \in \Theta$ where the parameter space $\Theta$ is compact.

3) Identification: $\mathbf{H}_{L,0} = \begin{bmatrix} \mathbf{z} - \mathbf{F}^{-1}(\mathbf{c};\phi) \\ \mathbf{y}_0 - \mathbf{L}(\mathbf{c};\phi) \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \Rightarrow \phi = \phi_0$ where $\mathbf{z} = \text{plim}\ \tilde{\mathbf{z}}$ and $\mathbf{y}_0 = \text{plim}\ \tilde{\mathbf{y}}$ .

4) Continuity: $\mathbf{F}^{-1}(\mathbf{c};\phi)$ and $\mathbf{L}(\mathbf{c};\phi)$ are continuous over $\Theta$.

5) $\mathbf{\Omega}_L$ is invertible.

6) $\phi_0$ is an interior point of $\Theta$.

7) $\mathbf{F}^{-1}(\mathbf{c};\phi)$ and $\mathbf{L}(\mathbf{c};\phi)$ are finite and continuously differentiable in a neighbourhood of $\phi_0$.

8) $\dfrac{\partial \mathbf{H}_L'(\phi_0)}{\partial \phi} \mathbf{W}_L \dfrac{\partial \mathbf{H}_L(\phi_0)}{\partial \phi'}$ is non-singular

Then $\hat{\phi}_L$ defined by (3.8) is a consistent estimator of $\phi_0$ and

$$\sqrt{T}\left(\hat{\phi}_L - \phi_0\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_L) \quad \text{with} \quad \mathbf{V}_L = \left(\frac{\partial \mathbf{H}_L'(\phi_0)}{\partial \phi} \mathbf{W}_L \frac{\partial \mathbf{H}_L(\phi_0)}{\partial \phi'}\right)^{-1}$$

***Proof*:** Assumption (1) implies that the cdf of the distribution is twice differentiable, and strictly monotonic with finite second order moment. This is equivalent to the assumption in Beach and Davidson (1983) or Cowell and Victoria-Feser (2002) guaranteeing $\tilde{\mathbf{z}} \to \mathbf{z}$ and $\tilde{\mathbf{y}} \to \mathbf{y}_0$ with covariance matrix $\mathbf{\Omega}_L$ in (3.7).

For any matrix $\tilde{\mathbf{W}}_L \xrightarrow{p} \mathbf{W}_L$ where $\mathbf{W}_L$ is a positive semidefinite matrix**,** assumptions 2 to 4, that is, compactness, identification, and continuity, plus an additional uniform convergence assumption, are standard sufficient conditions for consistency of $\hat{\phi}_L$ (see e.g., Newey and McFadden 1994, Theorem 2.1). For this problem, uniform convergence is satisfied given the other conditions. To show this we need to prove

$$\sup_{\phi\in\Theta}|Q(\phi)|\xrightarrow{p}0$$

$$\text{where } Q(\phi)=\begin{bmatrix}\tilde{\mathbf{z}}-\mathbf{F}^{-1}(\mathbf{c};\phi)\\ \tilde{\mathbf{y}}-\mathbf{L}(\mathbf{c};\phi)\end{bmatrix}'\tilde{\mathbf{W}}_L\begin{bmatrix}\tilde{\mathbf{z}}-\mathbf{F}^{-1}(\mathbf{c};\phi)\\ \tilde{\mathbf{y}}-\mathbf{L}(\mathbf{c};\phi)\end{bmatrix}-\begin{bmatrix}\mathbf{z}-\mathbf{F}^{-1}(\mathbf{c};\phi)\\ \mathbf{y}_0-\mathbf{L}(\mathbf{c};\phi)\end{bmatrix}'\mathbf{W}_L\begin{bmatrix}\mathbf{z}-\mathbf{F}^{-1}(\mathbf{c};\phi)\\ \mathbf{y}_0-\mathbf{L}(\mathbf{c};\phi)\end{bmatrix}$$

Performing the required multiplications and using the triangular inequality we can write:

$$|Q(\phi)|\le\left|\begin{bmatrix}\tilde{\mathbf{z}}-\mathbf{z}\\ \tilde{\mathbf{y}}-\mathbf{y}_0\end{bmatrix}'\tilde{\mathbf{W}}_L\begin{bmatrix}\tilde{\mathbf{z}}-\mathbf{z}\\ \tilde{\mathbf{y}}-\mathbf{y}_0\end{bmatrix}\right|+2\left|\begin{bmatrix}\mathbf{z}-\mathbf{F}^{-1}(\mathbf{c};\phi)\\ \mathbf{y}_0-\mathbf{L}(\mathbf{c};\phi)\end{bmatrix}'\tilde{\mathbf{W}}_L\begin{bmatrix}\tilde{\mathbf{z}}-\mathbf{z}\\ \tilde{\mathbf{y}}-\mathbf{y}_0\end{bmatrix}\right|+\left|\begin{bmatrix}\mathbf{F}^{-1}(\mathbf{c};\phi)\\ \mathbf{L}(\mathbf{c};\phi)\end{bmatrix}'\left(\tilde{\mathbf{W}}_L-\mathbf{W}_L\right)\begin{bmatrix}\mathbf{F}^{-1}(\mathbf{c};\phi)\\ \mathbf{L}(\mathbf{c};\phi)\end{bmatrix}\right|$$

Since each of the right hand side terms converges to zero irrespective of the value of $\phi$, we conclude that $\sup_{\phi\in\Theta}|Q(\phi)|\xrightarrow{p}0$ and thus uniform convergence is satisfied. Assumptions 4 and 5 are sufficient (through Slutsky's theorem) for showing $\mathbf{W}_L(\hat{\phi})\xrightarrow{p}\mathbf{W}_L$. Assumptions 6-8 plus $\sup_{\phi\in N(\phi_0)}\left\|\partial\mathbf{H}_L/\partial\phi-\partial\mathbf{H}_{L,0}/\partial\phi\right\|\to 0$ are sufficient conditions for asymptotic normality of a minimum distance estimator with covariance matrix as specified in (3.9) (see Newey and McFadden 1994, Theorem 3.20). The latter assumption is automatically satisfied because

$$\frac{\partial\mathbf{H}_L}{\partial\phi}-\frac{\partial\mathbf{H}_{L,0}}{\partial\phi}=\begin{bmatrix}\dfrac{\partial\mathbf{F}^{-1}}{\partial\phi}-\dfrac{\partial\mathbf{F}^{-1}}{\partial\phi}\\ \dfrac{\partial\mathbf{L}}{\partial\phi}-\dfrac{\partial\mathbf{L}}{\partial\phi}\end{bmatrix}=\mathbf{0}\text{ for any }\phi\in\Theta.$$

Similar theorems and proofs can be given for the cases where we only have the first $N-1$ moments (i.e., the quantile moments) or the last $N$ moments (i.e., the Lorenz curve moments).

**iii. $\mathbf{W}_{L,11},\mathbf{W}_{L,12}$ and $\mathbf{W}_{L,22}$ for *DGP 1* are tri-diagonal**

Denote

$$\mathbf{W}_L = \begin{pmatrix} \mathbf{W}_{L,11} & \mathbf{W}_{L,12} \\ \mathbf{W}_{L,21} & \mathbf{W}_{L,22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Omega}_{L,11} & \boldsymbol{\Omega}_{L,12} \\ \boldsymbol{\Omega}_{L,21} & \boldsymbol{\Omega}_{L,22} \end{pmatrix}^{-1}$$

Our aim is to prove that $\mathbf{W}_{L,11}$, $\mathbf{W}_{L,22}$ and $\mathbf{W}_{L,12}$ are tri-diagonal.

***Lemma-3***: For any invertible matrix $\mathbf{A}$ of the following form, $\mathbf{A}^{-1}$ is tri-diagonal

$$\mathbf{A} = \begin{pmatrix} a_1b_1 & a_1b_2 & \cdots & a_1b_N \\ a_1b_2 & a_2b_2 & \cdots & a_2b_N \\ \vdots & \vdots & \ddots & \vdots \\ a_1b_N & a_2b_N & \cdots & a_Nb_N \end{pmatrix}$$

This lemma can be proved by applying Cramer's rule for inverting matrices. The minors for elements not on the tri-diagonal positions are matrices where at least one of the columns is a multiple of another column and therefore has determinant zero, proving the lemma.

***Proof of Proposition 2*****:** Using the formula for inverse of partitioned matrices we can write

$$\mathbf{W}_L = \begin{pmatrix} \boldsymbol{\Omega}_{L,11}^{-1} + \boldsymbol{\Omega}_{L,11}^{-1}\boldsymbol{\Omega}_{L,12}(\boldsymbol{\Omega}_{L,22} - \boldsymbol{\Omega}_{L,21}\boldsymbol{\Omega}_{L,11}^{-1}\boldsymbol{\Omega}_{L,12})^{-1}\boldsymbol{\Omega}_{L,21}\boldsymbol{\Omega}_{L,11}^{-1} & -\boldsymbol{\Omega}_{L,11}^{-1}\boldsymbol{\Omega}_{L,12}(\boldsymbol{\Omega}_{L,22} - \boldsymbol{\Omega}_{L,21}\boldsymbol{\Omega}_{L,11}^{-1}\boldsymbol{\Omega}_{L,12})^{-1} \\ -(\boldsymbol{\Omega}_{L,22} - \boldsymbol{\Omega}_{L,21}\boldsymbol{\Omega}_{L,11}^{-1}\boldsymbol{\Omega}_{L,12})^{-1}\boldsymbol{\Omega}_{L,21}\boldsymbol{\Omega}_{L,11}^{-1} & (\boldsymbol{\Omega}_{L,22} - \boldsymbol{\Omega}_{L,21}\boldsymbol{\Omega}_{L,11}^{-1}\boldsymbol{\Omega}_{L,12})^{-1} \end{pmatrix}$$

Note that $\boldsymbol{\Omega}_{L,11}$ is in the form of matrix $\mathbf{A}$ and therefore its inverse is tri-diagonal. Using this fact and direct multiplication it can be shown that $(\boldsymbol{\Omega}_{L,22} - \boldsymbol{\Omega}_{L,21}\boldsymbol{\Omega}_{L,11}^{-1}\boldsymbol{\Omega}_{L,12})$ is in the form of $\mathbf{A}$ and hence its inverse is also tri-diagonal. Again, using this result and direct multiplication, it can be shown that the other blocks are tri-diagonal as well.

### iv. Estimated parameters have smaller asymptotic variance when z is observed

First consider the model developed for *DGP* 2 and define $\mathbf{G}_1$ as a $(2N-1)\times(N-1)$ matrix of derivatives of the moment conditions with respect to $\mathbf{z}$ and $\mathbf{G}_2$ as a $(2N-1)\times d$ matrix of derivatives of the moment conditions with respect to the parameters of the Lorenz curve $\boldsymbol{\phi}$.

If $\mathbf{z}$ is unknown the asymptotic covariance matrix for $\hat{\boldsymbol{\theta}}' = \left(\hat{\mathbf{z}}', \hat{\boldsymbol{\phi}}_m'\right)$ is

$$\text{cov}\left(\hat{\mathbf{z}},\hat{\phi}_m\right)=\frac{1}{T}\left(\begin{bmatrix}\mathbf{G}_1'\\ \mathbf{G}_2'\end{bmatrix}\mathbf{W}_m\begin{bmatrix}\mathbf{G}_1 & \mathbf{G}_2\end{bmatrix}\right)^{-1}=\frac{1}{T}\begin{pmatrix}\mathbf{G}_1'\mathbf{W}_m\mathbf{G}_1 & \mathbf{G}_1'\mathbf{W}_m\mathbf{G}_2\\ \mathbf{G}_2'\mathbf{W}_m\mathbf{G}_1 & \mathbf{G}_2'\mathbf{W}_m\mathbf{G}_2\end{pmatrix}^{-1}$$

from which we obtain

$$\text{cov}\left(\hat{\phi}_m\right)=\frac{1}{T}\left(\mathbf{G}_2'\mathbf{W}_m\mathbf{G}_2-\mathbf{G}_2'\mathbf{W}_m\mathbf{G}_1\left(\mathbf{G}_1'\mathbf{W}_m\mathbf{G}_1\right)^{-1}\mathbf{G}_1'\mathbf{W}_m\mathbf{G}_2\right)^{-1}$$

If $\mathbf{z}$ is known, the asymptotic covariance matrix for the estimator for $\phi$, call it $\hat{\phi}_z$, is

$$\text{cov}\left(\hat{\phi}_z\right)=\frac{1}{T}\left(\mathbf{G}_2'\mathbf{W}_m\mathbf{G}_2\right)^{-1}$$

Now, $\text{cov}\left(\hat{\phi}_m\right)-\text{cov}\left(\hat{\phi}_z\right)$ will be positive definite if

$$\left[\text{cov}\left(\hat{\phi}_z\right)\right]^{-1}-\left[\text{cov}\left(\hat{\phi}_m\right)\right]^{-1}=T\,\mathbf{G}_2'\mathbf{W}_m\mathbf{G}_1\left(\mathbf{G}_1'\mathbf{W}_m\mathbf{G}_1\right)^{-1}\mathbf{G}_1'\mathbf{W}_m\mathbf{G}_2$$

is positive definite. Since $\mathbf{W}_m$ is positive definite and $\mathbf{G}_1$ and $\mathbf{G}_2$ are of full rank from the assumed regularity conditions, $\mathbf{G}_2'\mathbf{W}_m\mathbf{G}_1\left(\mathbf{G}_1'\mathbf{W}_m\mathbf{G}_1\right)^{-1}\mathbf{G}_1'\mathbf{W}_m\mathbf{G}_2$ will be positive definite, and the claim is proved.

The above result together with Proposition 7 proves a similar result for *DGP 1*. That is, for *DGP 1* the estimator from a model with observed group bounds is more efficient than that from a model with unknown group bounds.

## Appendix 2

### i. Proposition 3: asymptotic distribution for $\mathbf{H}_m$

The result

$$\sqrt{T}\mathbf{H}_m=\sqrt{T}\begin{bmatrix}\tilde{\mathbf{c}}-\mathbf{F}(\mathbf{z};\phi)\\ \tilde{\mathbf{y}}_m-\mathbf{m}(\mathbf{z};\phi)\end{bmatrix}\rightarrow N(\mathbf{0},\mathbf{\Omega}_m)$$

follows from the Lindeberg-Levy central limit theorem and is well known in the literature, at least for the separate components $\sqrt{T}\left(\tilde{\mathbf{c}}-\mathbf{F}(\mathbf{z};\phi)\right)$ and $\sqrt{T}\left(\tilde{\mathbf{y}}_m-\mathbf{m}(\mathbf{z};\phi)\right)$. The only requirement is for the distribution to have finite first and second order moments. Here we use influence functions to derive the complete set of elements in the covariance matrix $\mathbf{\Omega}_m$.

***Lemma 4***: Influence function for cumulative proportions and income are

$$IF(y;F_i) = g(y \le z_i) - F_i$$
$$IF(y;m_i) = yg(y \le z_i) - m_i \tag{A2.1}$$

This can be proved by applying the result that, if $\tau$ is a linear functional, that is, $\tau = \int a(y) f(y)dy$, then $IF(y;\tau) = a(y) - \tau$. Defining $\tau = \int g(y) f(y)dy$ for $F_i$ and $\tau = \int yg(y) f(y)dy$ for $m_i$ proves the lemma.

Now, using the relationship between influence functions and asymptotic covariances we can derive the elements of the covariance matrix $\mathbf{\Omega}_m$ given in (4.4) as follows.

$$\mathrm{var}\left(\sqrt{T}\tilde{c}_i\right) = \int IF(y;F_i)^2 f(y)dy = \int [g(y \le z_i) - F_i]^2 f(y)dy = F_i - F_i^2 - F_i^2 + F_i^2 = F_i(1-F_i)$$

$$\mathrm{var}\left(\sqrt{T}\tilde{y}_{m,i}\right) = \int IF(y;m_i)^2 f(y)dy = \int [yg(y \le z_i) - m_i]^2 f(y)dy = m_i^{(2)} - m_i^2 - m_i^2 + m_i^2 = m_i^{(2)} - m_i^2$$

$$\mathrm{cov}\left(\sqrt{T}\tilde{c}_i, \sqrt{T}\tilde{y}_{m,i}\right) = \int IF(y;F_i) IF(y;m_i) f(y)dy = \int [g(y \le z_i) - F_i][yg(y \le z_i) - m_i] f(y)dy$$
$$= m_i - F_i m_i - F_i m_i + F_i\ m_i = m_i(1-F_i)$$

Also, for every $i$ and $j$ such that $i \le j$

$$\mathrm{cov}\left(\sqrt{T}\tilde{c}_i, \sqrt{T}\tilde{c}_j\right) = \int IF(y;F_i) IF(y;F_j) f(y)dy = \int [g(y \le z_i) - F_i][g(y \le z_j) - F_j] f(y)dy$$
$$= F_i - F_i F_j - F_i F_j + F_i F_j = F_i(1-F_j)$$

$$\mathrm{cov}\left(\sqrt{T}\tilde{y}_{m,i}, \sqrt{T}\tilde{y}_{m,j}\right) = \int IF(y;m_i) IF(y;m_j) f(y)dy = \int [yg(y \le z_i) - m_i][yg(y \le z_j) - m_j] f(y)dy$$
$$= m_i^{(2)} - m_i m_j - m_i m_j + m_i m_j = m_i^{(2)} - m_i m_j$$

$$\mathrm{cov}\left(\sqrt{T}\tilde{c}_i, \sqrt{T}\tilde{y}_{m,j}\right) = \int IF(y;F_i) IF(y;m_j) f(y)dy = \int [g(y \le z_i) - F_i][yg(y \le z_j) - m_j] f(y)dy$$
$$= m_i - F_i m_j - F_i m_j + F_i\ m_j = m_i - F_i m_j$$

**ii. Regularity conditions and proof of Proposition 4**

Suppose we have a random sample of size *T* from a population with cdf $F(y, \phi_0)$, with finite mean and variance, and that the following assumptions are valid.

1) Compactness: $\boldsymbol{\theta}_0 \in \Theta$ and the parameter space $\Theta$ is compact.

2) Identification: $\begin{bmatrix} \mathbf{c}-\mathbf{F}(\boldsymbol{\theta}) \\ \mathbf{y}_0-\mathbf{m}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \Rightarrow \boldsymbol{\theta} = \boldsymbol{\theta}_0$ where $\mathbf{c} = E(\tilde{\mathbf{c}})$ and $\mathbf{y}_0 = E(\tilde{\mathbf{y}})$.

3) Finiteness and Continuity: $\mathbf{F}(\boldsymbol{\theta})$ and $\mathbf{m}(\boldsymbol{\theta})$ are finite and continuous in $\boldsymbol{\theta} \in \Theta$.

4) $\mathbf{m}^{(2)}(\boldsymbol{\theta}_0) < \infty$ and the elements of $\mathbf{m}^{(2)}(\boldsymbol{\theta})$ are continuous for all $\boldsymbol{\theta} \in \Theta$.

5) $\boldsymbol{\Omega}_m$ is invertible.

6) $\boldsymbol{\theta}_0$ is an interior point of $\Theta$.

7) $\mathbf{F}(\boldsymbol{\theta})$ and $\mathbf{m}(\boldsymbol{\theta})$ are finite and continuously differentiable in neighbourhood $N(\boldsymbol{\theta}_0)$.

8) $\dfrac{\partial \mathbf{H}'_m(\tilde{\mathbf{c}},\tilde{\mathbf{y}};\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \mathbf{W}_m \dfrac{\partial \mathbf{H}_m(\tilde{\mathbf{c}},\tilde{\mathbf{y}};\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}$ is non-singular.

Then, $\hat{\boldsymbol{\theta}}$ obtained from (4.5) is a consistent estimator of $\boldsymbol{\theta}_0$ and

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right) \xrightarrow{d} N(\mathbf{0},\mathbf{V}_m) \quad \text{with} \quad \mathbf{V}_m = \left( \frac{\partial \mathbf{H}'_m}{\partial \boldsymbol{\theta}} \mathbf{W}_m \frac{\partial \mathbf{H}_m}{\partial \boldsymbol{\theta}'} \right)^{-1}$$

***Proof***: For any matrix $\tilde{\mathbf{W}}_m \xrightarrow{p} \mathbf{W}_m$ where $\mathbf{W}_m$ is a positive semidefinite matrix, assumptions 1-3 i.e. compactness, identification and continuity plus a uniform convergence assumption are standard sufficient conditions for consistency of $\hat{\boldsymbol{\theta}}$ (see e.g., Newey and McFadden 1994, theorem 2.1). Uniform convergence is also satisfied given assumptions 1-3. To show this, we use theorem 2.6 in Newey and McFadden (1994). From this theorem, uniform convergence is satisfied if we can show

$$E[\sup_{\theta \in \Theta} \|\mathbf{h}(y;\boldsymbol{\theta})\|] < \infty$$

where $\mathbf{h}(y;\boldsymbol{\theta})$ is a $(2N-1)$ dimensional vector with elements $g_i(y \le z) - F_i(\boldsymbol{\theta})$ for $i = 1,\ldots,N-1$, and $y g_i(y \le z) - m_i(\boldsymbol{\theta})$ for $i = N,\ldots,2N-1$. For some $j \in \{1,\ldots,N-1\}$, and $j' \in \{N,\ldots,2N-1\}$, we can write

$$\|\mathbf{h}(y;\boldsymbol{\theta})\| \le \max_i \sqrt{2N-1}\,|h_i(y;\boldsymbol{\theta})| = \begin{cases} \text{either } \sqrt{2N-1}\,|g_j(y \le z) - F_j(\boldsymbol{\theta})| \le \sqrt{2N-1}\left(1+F_j(\boldsymbol{\theta})\right) \\ \text{or } \quad \sqrt{2N-1}\,|y g_{j'}(y \le z) - m_{j'}(\boldsymbol{\theta})| \le \sqrt{2N-1}\left(y + m_{j'}(\boldsymbol{\theta})\right) \end{cases}$$

The first inequality is a well-known result regarding norms. Note that $E[\sup_{\theta\in\Theta}\{1+F_j(\boldsymbol{\theta})\}] = 1+F_j(\boldsymbol{\theta}^*)$ and $E[\sup_{\theta\in\Theta}\{y+m_{j'}(\boldsymbol{\theta})\}] = m + m_{j'}(\boldsymbol{\theta}^{**})$ for some $\boldsymbol{\theta}^*, \boldsymbol{\theta}^{**} \in \Theta$. $F_j(\boldsymbol{\theta})$ and $m_{j'}(\boldsymbol{\theta})$ are assumed finite for all $\boldsymbol{\theta} \in \Theta$ and thus uniform convergence is satisfied. Assumption 4 is required to show that $\sqrt{T}\begin{bmatrix} \tilde{\mathbf{c}} - \mathbf{F} \\ \tilde{\mathbf{y}} - \mathbf{m} \end{bmatrix} \to N(\mathbf{0}, \boldsymbol{\Omega}_m)$ using the Lindeberg-Levy central limit theorem; due to continuity of $\mathbf{F}(\boldsymbol{\theta})$, $\mathbf{m}(\boldsymbol{\theta})$ and $\mathbf{m}^{(2)}(\boldsymbol{\theta})$, we can write $\mathbf{W}_m(\tilde{\boldsymbol{\theta}}) \xrightarrow{p} \mathbf{W}_m(\boldsymbol{\theta}_0)$. Assumptions 5-8 plus $E\left\{\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} \|\partial\mathbf{h}/\partial\boldsymbol{\theta}'\|\right\} < \infty$ are sufficient conditions for asymptotic normality of the GMM estimator with covariance matrix specified in (4.6) (see e.g. Newey and McFadden 1994, theorem 3.4); $E\left\{\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} \|\partial\mathbf{h}/\partial\boldsymbol{\theta}'\|\right\}$ is finite because $\dfrac{\partial\mathbf{h}}{\partial\boldsymbol{\theta}'} = \begin{bmatrix} \partial\mathbf{F}/\partial\boldsymbol{\theta}' \\ \partial\mathbf{m}/\partial\boldsymbol{\theta}' \end{bmatrix}$ is finite and not dependent on $y$.

A similar theorem and proof can be given for the case where the group bounds $\mathbf{z}$ are known by replacing $\boldsymbol{\theta}$ with $\boldsymbol{\phi}$.

### iii. Derivation of the optimal weight matrix $\mathbf{W}_m$

Obtaining the weight matrix given in Proposition 5 requires inverting (4.4). Hajargasht et al. (2012) have derived the weight matrix for a situation where population proportions and income are defined non-cumulatively. To use their result to derive the weight matrix for the cumulative population proportions and income considered here, we begin by defining the $(N \times N)$ matrix $\mathbf{A}_N$ and the $[(2N-1)\times(2N-1)]$ matrix $\mathbf{B}$ as follows.

$$\mathbf{A}_N = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} \mathbf{A}_{N-1} & \mathbf{0}_{N-1,N} \\ \mathbf{0}_{N,N-1} & \mathbf{A}_N \end{bmatrix} \tag{A2.2}$$

Let $\mathbf{H}^*(\boldsymbol{\theta}) = \mathbf{B}\mathbf{H}_m(\boldsymbol{\theta})$. The relationship between the asymptotic covariance matrices for $\mathbf{H}^*(\boldsymbol{\theta})$ and $\mathbf{H}_m(\boldsymbol{\theta})$ is $\text{var}\left[T^{1/2}\mathbf{H}^*(\boldsymbol{\theta})\right] = \mathbf{B}\,\text{var}\left[T^{1/2}\mathbf{H}_m(\boldsymbol{\theta})\right]\mathbf{B}'$ from which we obtain

$$\text{var}\left[T^{1/2}\mathbf{H}_m(\boldsymbol{\theta})\right] = \mathbf{B}^{-1}\,\text{var}\left[T^{1/2}\mathbf{H}^*(\boldsymbol{\theta})\right]\mathbf{B}'^{-1} \tag{A2.3}$$

The required weight matrix is

$$\mathbf{W}_m = \left(\text{var}\left[T^{1/2}\mathbf{H}_m(\boldsymbol{\theta})\right]\right)^{-1} = \mathbf{B}'\left(\text{var}\left[T^{1/2}\mathbf{H}^*(\boldsymbol{\theta})\right]\right)^{-1}\mathbf{B} \tag{A2.4}$$

In Hajargasht et al. (2012) it is shown that

$$\left(\text{var}\left[T^{1/2}\mathbf{H}^*(\boldsymbol{\theta})\right]\right)^{-1} = \begin{bmatrix} D\left(\boldsymbol{\mu}_{-N}^{(2)}/\mathbf{v}_{-N}\right) + \left(\mu_N^{(2)}/v_N\right)\mathbf{j}_{N-1}\mathbf{j}'_{N-1} & -D\left(\boldsymbol{\mu}_{-N}/\mathbf{v}_{-N}\right) & \left(\mu_N/v_N\right)\mathbf{j}_{N-1} \\ -D\left(\boldsymbol{\mu}_{-N}/\mathbf{v}_{-N}\right) & D\left(\boldsymbol{\kappa}_{-N}/\mathbf{v}_{-N}\right) & \mathbf{0} \\ \left(\mu_N/v_N\right)\mathbf{j}'_{N-1} & \mathbf{0} & \kappa_N/v_N \end{bmatrix} \tag{A2.5}$$

where $\boldsymbol{\mu}'_{-N} = \left(\mu_1, \mu_2, \ldots, \mu_{N-1}\right)$, $\boldsymbol{\mu}'^{(2)}_{-N} = \left(\mu_1^{(2)}, \mu_2^{(2)}, \ldots, \mu_N^{(2)}\right)$, $\boldsymbol{\kappa}'_{-N} = \left(\kappa_1, \kappa_2, \ldots, \kappa_{N-1}\right)$, and $\mathbf{v}'_{-N} = \left(v_1, v_2, \ldots, v_{N-1}\right)$. Using (A2.2) and (A2.5) to carry out the matrix multiplication in (A2.4) yields the weight matrix given in Proposition 5.

## Appendix 3 Proof of proposition 7

*(i)* Using Lemma 2, and the matrix $\mathbf{A}$ defined in Proposition 6, it can be shown that

$$\begin{bmatrix} \dfrac{\partial \mathbf{F}'^{-1}}{\partial \boldsymbol{\phi}} & \dfrac{\partial \mathbf{L}'}{\partial \boldsymbol{\phi}} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial \mathbf{F}'}{\partial \boldsymbol{\phi}} & \dfrac{\partial \mathbf{m}'}{\partial \boldsymbol{\phi}} \end{bmatrix}\mathbf{A}'$$

From this result and that $\mathbf{W}_L = \mathbf{A}'^{-1}\mathbf{W}_m\mathbf{A}^{-1}$ from Proposition 6, we have

$$\text{cov}\left(\hat{\boldsymbol{\phi}}_L\right) = \frac{1}{T}\left(\begin{bmatrix} \dfrac{\partial \mathbf{F}'^{-1}}{\partial \boldsymbol{\phi}} & \dfrac{\partial \mathbf{L}'}{\partial \boldsymbol{\phi}} \end{bmatrix}\mathbf{W}_L(\boldsymbol{\phi})\begin{bmatrix} \partial \mathbf{F}^{-1}/\partial \boldsymbol{\phi}' \\ \partial \mathbf{L}/\partial \boldsymbol{\phi}' \end{bmatrix}\right)^{-1} = \frac{1}{T}\left(\begin{bmatrix} \dfrac{\partial \mathbf{F}'}{\partial \boldsymbol{\phi}} & \dfrac{\partial \mathbf{m}'}{\partial \boldsymbol{\phi}} \end{bmatrix}\mathbf{A}'\mathbf{A}'^{-1}\mathbf{W}_m\mathbf{A}^{-1}\mathbf{A}\begin{bmatrix} \partial \mathbf{F}/\partial \boldsymbol{\phi}' \\ \partial \mathbf{m}/\partial \boldsymbol{\phi}' \end{bmatrix}\right)^{-1}$$

$$= \frac{1}{T}\left(\begin{bmatrix} \dfrac{\partial \mathbf{F}'}{\partial \boldsymbol{\phi}} & \dfrac{\partial \mathbf{m}'}{\partial \boldsymbol{\phi}} \end{bmatrix}\mathbf{W}_m\begin{bmatrix} \partial \mathbf{F}/\partial \boldsymbol{\phi}' \\ \partial \mathbf{m}/\partial \boldsymbol{\phi}' \end{bmatrix}\right)^{-1} = \text{cov}(\hat{\boldsymbol{\phi}}_m)$$

*(ii)* First note that

$$\text{cov}(\hat{\boldsymbol{\phi}}_L) = \frac{1}{T}\left(\frac{\partial \mathbf{L}'}{\partial \boldsymbol{\phi}}\boldsymbol{\Omega}_{L,22}^{-1}(\boldsymbol{\phi})\frac{\partial \mathbf{L}}{\partial \boldsymbol{\phi}'}\right)^{-1}$$

Recall that $\partial \mathbf{H}'_m / \partial \boldsymbol{\phi} = \left( \partial \mathbf{F}' / \partial \boldsymbol{\phi} \quad \partial \mathbf{m}' / \partial \boldsymbol{\phi} \right)$. From Proposition 6 and Lemma 2(iii), and recognising that $\partial L_N / \partial \boldsymbol{\phi} = \partial m_N / \partial \boldsymbol{\phi}$, we can write $\partial \mathbf{L} / \partial \boldsymbol{\phi}' = \mathbf{B} \left( \partial \mathbf{H}_m / \partial \boldsymbol{\phi}' \right)$, and

$$\begin{aligned} \operatorname{cov}(\hat{\boldsymbol{\phi}}_L) &= \frac{1}{T} \left( \frac{\partial \mathbf{H}'_m}{\partial \boldsymbol{\phi}} \mathbf{B}' (\mathbf{B}\mathbf{B}')^{-1} \mathbf{B} [\mathbf{W}_m - \mathbf{W}_m \mathbf{C} (\mathbf{C}' \mathbf{W}_m \mathbf{C})^{-1} \mathbf{C}' \mathbf{W}_m ] \mathbf{B}' (\mathbf{B}\mathbf{B}')^{-1} \mathbf{B} \frac{\partial \mathbf{H}_m}{\partial \boldsymbol{\phi}'} \right)^{-1} \\ &= \frac{1}{T} \left( \frac{\partial \mathbf{H}'_m}{\partial \boldsymbol{\phi}} [\mathbf{W}_m - \mathbf{W}_m \mathbf{C} (\mathbf{C}' \mathbf{W}_m \mathbf{C})^{-1} \mathbf{C}' \mathbf{W}_m ] \frac{\partial \mathbf{H}_m}{\partial \boldsymbol{\phi}'} \right)^{-1} \end{aligned}$$

The last equality comes from a result established in Proposition 6, namely,

$$[\mathbf{I} - \mathbf{C} (\mathbf{C}' \mathbf{W}_m \mathbf{C})^{-1} \mathbf{C}' \mathbf{W}_m ] \, \mathbf{B}' (\mathbf{B}\mathbf{B}')^{-1} \mathbf{B} = [\mathbf{I} - \mathbf{C} (\mathbf{C}' \mathbf{W}_m \mathbf{C})^{-1} \mathbf{C}' \mathbf{W}_m ]$$

Using Proposition 4,

$$\operatorname{cov} \begin{pmatrix} \hat{\mathbf{z}} \\ \hat{\boldsymbol{\phi}}_m \end{pmatrix} = \frac{1}{T} \begin{pmatrix} \left( \partial \mathbf{H}'_m / \partial \mathbf{z} \right) \mathbf{W}_m \left( \partial \mathbf{H}_m / \partial \mathbf{z}' \right) & \left( \partial \mathbf{H}'_m / \partial \mathbf{z} \right) \mathbf{W}_m \left( \partial \mathbf{H}_m / \partial \boldsymbol{\phi}' \right) \\ \left( \partial \mathbf{H}'_m / \partial \boldsymbol{\phi} \right) \mathbf{W}_m \left( \partial \mathbf{H}_m / \partial \mathbf{z}' \right) & \left( \partial \mathbf{H}'_m / \partial \boldsymbol{\phi} \right) \mathbf{W}_m \left( \partial \mathbf{H}_m / \partial \boldsymbol{\phi}' \right) \end{pmatrix}^{-1}$$

Using the formula for the inverse of a partitioned matrix, we have

$$\operatorname{cov}\left( \hat{\boldsymbol{\phi}}_m \right) = \frac{1}{T} \left( \frac{\partial \mathbf{H}'_m}{\partial \boldsymbol{\phi}} \mathbf{W}_m \frac{\partial \mathbf{H}_m}{\partial \boldsymbol{\phi}'} - \frac{\partial \mathbf{H}'_m}{\partial \boldsymbol{\phi}} \mathbf{W}_m \frac{\partial \mathbf{H}_m}{\partial \mathbf{z}'} \left( \frac{\partial \mathbf{H}'_m}{\partial \mathbf{z}} \mathbf{W}_m \frac{\partial \mathbf{H}_m}{\partial \mathbf{z}'} \right)^{-1} \frac{\partial \mathbf{H}'_m}{\partial \mathbf{z}} \mathbf{W}_m \frac{\partial \mathbf{H}_m}{\partial \boldsymbol{\phi}'} \right)^{-1}$$

Now,

$$\frac{\partial \mathbf{H}'_m}{\partial \mathbf{z}} = \left[ D\left( \mathbf{f}(\mathbf{z}) \right) \quad D\left( \mathbf{z}.\mathbf{f}(\mathbf{z}) \right) \quad \mathbf{0} \right] = D\left( \mathbf{f}(\mathbf{z}) \right) \mathbf{C}'$$

and thus,

$$\begin{aligned} \operatorname{cov}\left( \hat{\boldsymbol{\phi}}_m \right) &= \frac{1}{T} \left( \frac{\partial \mathbf{H}'_m}{\partial \boldsymbol{\phi}} \left\{ \mathbf{W}_m - \mathbf{W}_m \mathbf{C} D\left( \mathbf{f}(\mathbf{z}) \right) \left[ D\left( \mathbf{f}(\mathbf{z}) \right) \mathbf{C}' \mathbf{W}_m \mathbf{C} D\left( \mathbf{f}(\mathbf{z}) \right) \right]^{-1} D\left( \mathbf{f}(\mathbf{z}) \right) \mathbf{C}' \mathbf{W}_m \right\} \frac{\partial \mathbf{H}_m}{\partial \boldsymbol{\phi}'} \right)^{-1} \\ &= \frac{1}{T} \left( \frac{\partial \mathbf{H}'_m}{\partial \boldsymbol{\phi}} [\mathbf{W}_m - \mathbf{W}_m \mathbf{C} (\mathbf{C}' \mathbf{W}_m \mathbf{C})^{-1} \mathbf{C}' \mathbf{W}_m ] \frac{\partial \mathbf{H}_m}{\partial \boldsymbol{\phi}'} \right)^{-1} \end{aligned}$$

which proves the proposition.

## Appendix 4 Inference for standard Lorenz curves

The above results can be used to provide inference based on the standard Lorenz curve. We can define the moment conditions for the standard Lorenz curve by

$$DGP\ 1:\quad \underset{(2N-2)\times 1}{\mathbf{P}_L} = \begin{bmatrix} \tilde{z}_1 - F^{-1}(c_1) \\ \vdots \\ \tilde{z}_{N-1} - F^{-1}(c_{N-1}) \\ \tilde{y}_{L,1}/\tilde{y}_{L,N} - L_1/m_N \\ \vdots \\ \tilde{y}_{L,N-1}/\tilde{y}_{L,N} - L_{N-1}/m_N \end{bmatrix} \qquad DGP\ 2:\quad \underset{(2N-2)\times 1}{\mathbf{P}_m} = \begin{bmatrix} c_1 - F_1 \\ \vdots \\ c_{N-1} - F_{N-1} \\ \tilde{y}_{m,1}/\tilde{y}_{m,N} - m_1/m_N \\ \vdots \\ \tilde{y}_{m,N-1}/\tilde{y}_{m,N} - m_{N-1}/m_N \end{bmatrix}$$

For *DGP 1*, let

$$\boldsymbol{\eta} = \begin{bmatrix} \tilde{z}_1 \\ \vdots \\ \tilde{z}_{N-1} \\ \tilde{y}_{L,1}/\tilde{y}_{L,N} \\ \vdots \\ \tilde{y}_{L,N-1}/\tilde{y}_{L,N} \end{bmatrix} \quad \text{and} \quad \tilde{\boldsymbol{\eta}} = \begin{bmatrix} \tilde{z}_1 \\ \vdots \\ \tilde{z}_{N-1} \\ \tilde{y}_{L,1} \\ \vdots \\ \tilde{y}_{L,N} \end{bmatrix} \quad \text{so that} \quad \boldsymbol{\eta} = \begin{bmatrix} \tilde{\eta}_1 \\ \vdots \\ \tilde{\eta}_{N-1} \\ \tilde{\eta}_N/\tilde{\eta}_{2N-1} \\ \vdots \\ \tilde{\eta}_{2N-2}/\tilde{\eta}_{2N-1} \end{bmatrix} \tag{A4.1}$$

Then, using the delta method, the covariance matrix of the limiting distribution can be written as

$$\text{cov}\left[T^{1/2}\mathbf{P}_L\right] = \text{plim}\left(\frac{\partial\boldsymbol{\eta}}{\partial\tilde{\boldsymbol{\eta}}'}\right)\text{cov}\left(T^{1/2}\tilde{\boldsymbol{\eta}}\right)\text{plim}\left(\frac{\partial\boldsymbol{\eta}'}{\partial\tilde{\boldsymbol{\eta}}}\right) \tag{A4.2}$$

Differentiating and taking probability limits, yields

$$\text{plim}\left(\frac{\partial\boldsymbol{\eta}}{\partial\tilde{\boldsymbol{\eta}}'}\right) = \begin{pmatrix} \mathbf{I}_{N-1} & \mathbf{0}_{N-1\times N} \\ \mathbf{0}_{N-1} & \begin{bmatrix} \dfrac{\mathbf{I}_{N-1}}{m_N} & -\dfrac{\mathbf{m}_{-N}}{m_N^2} \end{bmatrix} \end{pmatrix}_{(2N-2)\times(2N-1)} \tag{A4.3}$$

A similar result can be obtained for *DGP 2*. Nothing that $\text{cov}\left(T^{1/2}\tilde{\boldsymbol{\eta}}\right)$ is given in (3.7), the required asymptotic variance can be computed from (A4.2) and (A4.3). For the weight matrix in minimum distance estimation $\text{cov}\left[T^{1/2}\mathbf{P}_L\right]$ needs to be inverted.

## Appendix 5 Lorenz curves with individual data

**Proof of Proposition 8:** In Proposition 1 we were concerned with the consistency and asymptotic normality of the minimum distance estimator that considers both the quantile function and the Lorenz curve. In Proposition 8 we focus on only the conditions for the quantile function. If, for the moment, we consider a fixed number of groups, the proof for consistency of the estimator in the first line in (8.1) and its asymptotic normality in (8.2) is similar to that of Proposition 1. The following assumptions are required.

1) $F^{-1}(c;\phi_0)$ is differentiable with respect to $c$ and $0 < \partial F^{-1}(c;\phi_0)/\partial c < \infty$ over (0,1).

2) Compactness: $\phi_0 \in \Theta$ where the parameter space $\Theta$ is compact.

3) Identification: $\mathbf{z} - \mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi}) = \mathbf{0} \Rightarrow \boldsymbol{\phi} = \boldsymbol{\phi}_0$ where $\mathbf{z} = \text{plim}\,\tilde{\mathbf{z}}$.

4) Continuity: $\mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi})$ is continuous over $\Theta$.

5) $\boldsymbol{\phi}_0$ is an interior point of $\Theta$

6) $\mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi})$ is finite and continuously differentiable in a neighbourhood of $\boldsymbol{\phi}_0$.

7) $\dfrac{\partial \mathbf{F}'^{-1}(\mathbf{c};\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}} \mathbf{W}_{L,11} \dfrac{\partial \mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}'}$ is non-singular.

The proposition is also true when the number of groups increases with the sample size, as is the case with individual data, but some of the underlying assumptions need to be checked to ensure they can be satisfied. The main assumption to check for consistency is the uniform convergence assumption. In Proposition 1 where we have a finite number of groups it is implied by other assumptions. Here we show that it is still implied by the other assumptions, even for the individual observations case.

Let $Q(\boldsymbol{\phi})$ be the objective function; we must find a $Q_0(\boldsymbol{\phi})$ where $\sup_{\boldsymbol{\phi}\in\Theta} |Q(\boldsymbol{\phi}) - Q_0(\boldsymbol{\phi})| \xrightarrow{p} 0$ and $\boldsymbol{\phi}_0$ is the unique minimiser of $Q_0(\boldsymbol{\phi})$. Define $Q_0(\boldsymbol{\phi})$ as

$Q_0(\boldsymbol{\phi}) = \left(\mathbf{z} - \mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi})\right)' \mathbf{W}_{L,11}\left(\mathbf{z} - \mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi})\right) + 1$ . First note that this function is minimised at $\boldsymbol{\phi}_0$ . Second, using the triangular inequality we can write

$$\left|Q(\boldsymbol{\phi}) - Q_0(\boldsymbol{\phi}_0)\right| \leq \left|\left(\tilde{\mathbf{z}} - \mathbf{z}\right)' \tilde{\mathbf{W}}_{L,11}\left(\tilde{\mathbf{z}} - \mathbf{z}\right) - 1\right|$$
$$+ 2\left|\left(\mathbf{z} - \mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi})\right)' \tilde{\mathbf{W}}_{L,11}\left(\tilde{\mathbf{z}} - \mathbf{z}\right)\right| + \left|\mathbf{F}'^{-1}(\mathbf{c};\boldsymbol{\phi})\left(\tilde{\mathbf{W}}_{L,11} - \mathbf{W}_{L,11}\right)\mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi})\right|$$

Each term on the right-hand side converges to zero for any $\boldsymbol{\phi} \in \Theta$ when the sample size goes to infinity. For the first term, note that, for every $N$ , $\left(\tilde{\mathbf{z}} - \mathbf{z}\right)' \tilde{\mathbf{W}}_{L,11}\left(\tilde{\mathbf{z}} - \mathbf{z}\right) \to \chi^2(N-1)/T$ and when $N$ grows we have $\chi^2(N-1)/T \to 1$. For the second term note that again, for every $N$ , we have $\left(\tilde{\mathbf{z}} - \mathbf{z}\right) \to 0$ and lastly we assume $\left(\tilde{\mathbf{W}}_{L,11} - \mathbf{W}_{L,11}\right) \to 0$ for every $N$ .

For asymptotic normality we start with the first order conditions:

$$\frac{\partial \mathbf{F}'^{-1}(\mathbf{c};\hat{\boldsymbol{\phi}})}{\partial \hat{\boldsymbol{\phi}}} \tilde{\mathbf{W}}_{L,11}\left(\tilde{\mathbf{z}} - \mathbf{F}^{-1}(\mathbf{c};\hat{\boldsymbol{\phi}})\right) = \mathbf{0}$$

Using a Taylor expansion we can write

$$-\frac{\partial \mathbf{F}'^{-1}(\mathbf{c};\hat{\boldsymbol{\phi}})}{\partial \hat{\boldsymbol{\phi}}} \tilde{\mathbf{W}}_{L,11}\left(\tilde{\mathbf{z}} - \mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi}_0) + \frac{\partial \mathbf{F}^{-1}(\mathbf{c};\tilde{\boldsymbol{\phi}})}{\partial \tilde{\boldsymbol{\phi}}'}\left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0\right)\right) = \mathbf{0}$$

where $\tilde{\boldsymbol{\phi}}$ is a mean value, i.e., $\tilde{\boldsymbol{\phi}} = \lambda\boldsymbol{\phi}_0 + (1-\lambda)\hat{\boldsymbol{\phi}}$ for some $0 \leq \lambda \leq 1$. Solving in terms of $\left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0\right)$, we obtain

$$\sqrt{T}\left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0\right) = -\left(\frac{\partial \mathbf{F}'^{-1}(\mathbf{c};\hat{\boldsymbol{\phi}})}{\partial \hat{\boldsymbol{\phi}}} \tilde{\mathbf{W}}_{L,11} \frac{\partial \mathbf{F}^{-1}(\mathbf{c};\tilde{\boldsymbol{\phi}})}{\partial \tilde{\boldsymbol{\phi}}'}\right)^{-1} \frac{\partial \mathbf{F}'^{-1}(\mathbf{c};\tilde{\boldsymbol{\phi}})}{\partial \tilde{\boldsymbol{\phi}}} \tilde{\mathbf{W}}_{L,11}\sqrt{T}\left(\tilde{\mathbf{z}} - \mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi}_0)\right)$$

Let $\mathbf{V} = \left[\frac{\partial \mathbf{F}'^{-1}(\mathbf{c};\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}_0} \mathbf{W}_{L,11} \frac{\partial \mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi}_0)}{\partial \boldsymbol{\phi}_0'}\right]^{-1}$. Because $\sqrt{T}\left(\tilde{\mathbf{z}} - \mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi}_0)\right) \to N(\mathbf{0}, \boldsymbol{\Omega}_{L,11})$ for every $N$ , we can write

$$\sqrt{T}\left(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0\right)\rightarrow N\left\{\mathbf{0},\mathbf{V}\left[\frac{\partial\mathbf{F}'^{-1}(\mathbf{c};\boldsymbol{\phi}_0)}{\partial\boldsymbol{\phi}_0}\mathbf{W}_{L,11}\boldsymbol{\Omega}_{L,11}\mathbf{W}_{L,11}\frac{\partial\mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi}_0)}{\partial\boldsymbol{\phi}_0'}\right]\mathbf{V}\right\}$$

To obtain this result, we have used standard probability limit rules, uniform continuity of $\frac{\partial\mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi})}{\partial\boldsymbol{\phi}}$ over $\Theta$, and the Slutsky theorem. Note further that that we have $\mathbf{W}_{L,11}\boldsymbol{\Omega}_{L,11}=\mathbf{I}$ as long as $\boldsymbol{\Omega}_{L,11}$ is an invertible matrix for every $N$. We see below that this is in fact the case. Thus, we can write $\sqrt{T}\left(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0\right)\rightarrow N\left(\mathbf{0},\mathbf{V}\right)$.

To derive the weight matrix in (8.3), we note, from (3.2) that

$$\mathbf{W}_{L,11}=\Omega_{L,11}^{-1}=\begin{bmatrix}\frac{c_1(1-c_1)}{f(z_1)^2} & \cdots & \frac{c_1(1-c_{N-1})}{f(z_1)f(z_{N-1})}\\ \vdots & \ddots & \vdots\\ \frac{c_1(1-c_{N-1})}{f(z_1)f(z_{N-1})} & \cdots & \frac{c_{N-1}(1-c_{N-1})}{f(z_{N-1})^2}\end{bmatrix}^{-1}=D[\mathbf{f}(\mathbf{z})]\,\mathbf{A}\,D[\mathbf{f}(\mathbf{z})]$$

where, using direct multiplication, we can show that

$$\mathbf{A}=\begin{bmatrix}\frac{c_2-c_0}{(c_1-c_0)(c_2-c_1)} & \frac{-1}{(c_2-c_1)} & & \\ \frac{-1}{(c_2-c_1)} & \ddots & \ddots & \\ & \ddots & \ddots & \frac{-1}{(c_N-c_{N-1})}\\ & & \frac{-1}{(c_N-c_{N-1})} & \frac{c_N-c_{N-2}}{(c_N-c_{N-1})(c_{N-1}-c_{N-2})}\end{bmatrix}=N\begin{bmatrix}2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1\\ & & -1 & 2\end{bmatrix}$$

The right hand side is true if $\mathbf{c}'=\left(1/N,2/N,\ldots,(N-1)/N\right)$, $c_0=0$ and $c_N=1$ as is the case with individual data. Noting that $f(z_i;\boldsymbol{\phi})=\frac{1}{\partial F^{-1}(c_i,\boldsymbol{\phi})/\partial c_i}$ proves (8.3).

***Proposition 9***

To prove Proposition 9 we provide a series of lemmas:

***Lemma 5:*** Let $\mathbf{H}_1(\boldsymbol{\theta})$ define a set of $N$ moment conditions, $\mathbf{W}_1$ be the corresponding optimal GMM weight matrix, and $\hat{\boldsymbol{\theta}}$ be the optimal GMM estimator, with asymptotic covariance matrix $\text{cov}\left(\hat{\boldsymbol{\theta}}\right) = \frac{1}{T}\left(\frac{\partial \mathbf{H}_1'}{\partial \boldsymbol{\theta}} \mathbf{W}_1 \frac{\partial \mathbf{H}_1}{\partial \boldsymbol{\theta}'}\right)^{-1}$. Suppose there is also another $M$ set of moment conditions $\mathbf{H}(\boldsymbol{\theta})$ where we can write $\mathbf{H}(\boldsymbol{\theta}) = \mathbf{H}_2\left(\mathbf{H}_1(\boldsymbol{\theta})\right)$. If $M = N$ and $\partial \mathbf{H}_2 / \partial \mathbf{H}_1'$ evaluated at $\boldsymbol{\theta}_0$ is invertible, then $\text{cov}\left(\hat{\boldsymbol{\theta}}\right) = \text{cov}\left(\hat{\hat{\boldsymbol{\theta}}}\right)$ where $\hat{\hat{\boldsymbol{\theta}}}$ is the optimal GMM estimator corresponding to the moment conditions $\mathbf{H}(\boldsymbol{\theta})$.

***Proof:*** First note that using minimum distance estimation theory and the delta method, the optimal weight matrix and the asymptotic covariance matrix for $\hat{\hat{\boldsymbol{\theta}}}$ can be written as

$$\mathbf{W} = \left(\text{cov}[\mathbf{H}(\boldsymbol{\theta})]\right)^{-1} = \left(\frac{\partial \mathbf{H}_2}{\partial \mathbf{H}_1'} \text{cov}(\mathbf{H}_1) \frac{\partial \mathbf{H}_2'}{\partial \mathbf{H}_1}\right)^{-1} = \left(\frac{\partial \mathbf{H}_2}{\partial \mathbf{H}_1'} \mathbf{W}_1^{-1} \frac{\partial \mathbf{H}_2'}{\partial \mathbf{H}_1}\right)^{-1}$$

$$\text{cov}\left(\hat{\hat{\boldsymbol{\theta}}}\right) = \frac{1}{T}\left(\frac{\partial \mathbf{H}'}{\partial \boldsymbol{\theta}} \mathbf{W} \frac{\partial \mathbf{H}}{\partial \boldsymbol{\theta}'}\right)^{-1}$$

Using the chain rule we can write

$$\text{cov}\left(\hat{\hat{\boldsymbol{\theta}}}\right) = \frac{1}{T}\left(\frac{\partial \mathbf{H}_1'}{\partial \boldsymbol{\theta}} \frac{\partial \mathbf{H}_2'}{\partial \mathbf{H}_1}\left(\frac{\partial \mathbf{H}_2}{\partial \mathbf{H}_1'} \mathbf{W}_1^{-1} \frac{\partial \mathbf{H}_2'}{\partial \mathbf{H}_1}\right)^{-1} \frac{\partial \mathbf{H}_2}{\partial \mathbf{H}_1'} \frac{\partial \mathbf{H}_1}{\partial \boldsymbol{\theta}'}\right)^{-1}$$

If $\partial \mathbf{H}_2 / \partial \mathbf{H}_1'$ is invertible, then it follows that that $\text{cov}\left(\hat{\boldsymbol{\theta}}\right) = \text{cov}\left(\hat{\hat{\boldsymbol{\theta}}}\right)$.

***Lemma 6*:** Consider the model in (4.5) with moment conditions $\mathbf{H}_m\left(\boldsymbol{\phi}\right) = \begin{bmatrix} \tilde{\mathbf{c}} - \mathbf{F}(\mathbf{z};\boldsymbol{\phi}) \\ \tilde{\mathbf{y}}_m - \mathbf{m}(\mathbf{z};\boldsymbol{\phi}) \end{bmatrix}$ where the bounds are defined by $z_i = F^{-1}(i/N;\boldsymbol{\phi})$. If in the variance formula $N = T \to \infty$, the covariance matrix of the GMM estimator for $\boldsymbol{\phi}$ approaches the Cramer-Rao lower bound, i.e.,

$$\lim_{T\to\infty}\left(\frac{\partial \mathbf{H}_m'}{\partial \boldsymbol{\phi}} \mathbf{W}_m \frac{\partial \mathbf{H}_m}{\partial \boldsymbol{\phi}'}\right)^{-1} = \left(E\left\{\frac{\partial \ln \mathbf{f}'}{\partial \boldsymbol{\phi}} \frac{\partial \ln \mathbf{f}}{\partial \boldsymbol{\phi}'}\right\}\right)^{-1}$$

***Proof*:** This Lemma is proved in several steps:

(i) The optimal GMM estimator for $\boldsymbol{\phi}$ from the moment conditions in Lemma 6 has the same asymptotic covariance as the GMM estimator for the following problem

$$\mathbf{H}_2(\boldsymbol{\phi}) = \begin{bmatrix} \breve{\mathbf{c}} - \boldsymbol{\kappa} \\ \overline{\mathbf{y}} - \breve{\boldsymbol{\mu}} \end{bmatrix} = \begin{bmatrix} \breve{\mathbf{c}} - \boldsymbol{\kappa} \\ \breve{\mathbf{y}}/\breve{\mathbf{c}} - \boldsymbol{\mu}/\boldsymbol{\kappa} \end{bmatrix}$$

where $\breve{y}_\ell = \tilde{y}_\ell - \tilde{y}_{\ell-1}$, $\breve{c}_\ell = \tilde{c}_\ell - \tilde{c}_{\ell-1}$, $\overline{\mathbf{y}} = \breve{\mathbf{y}}/\breve{\mathbf{c}}$, and $\breve{\boldsymbol{\mu}} = \boldsymbol{\mu}/\boldsymbol{\kappa}$.

(ii) Let $\hat{\boldsymbol{\phi}}$ be the optimal GMM estimator from using $\mathbf{H}_2(\boldsymbol{\phi})$. Its asymptotic covariance matrix is given by

$$\text{cov}(\hat{\boldsymbol{\phi}}) = \frac{1}{T}\boldsymbol{\Upsilon}^{-1} \quad \text{where} \quad [\boldsymbol{\Upsilon}]_{ij} = \sum_{\ell=1}^{N} \left\{ \frac{(\partial\kappa_\ell/\partial\phi_i)(\partial\kappa_\ell/\partial\phi_j)}{\kappa_\ell} + \frac{(\partial\breve{\mu}_\ell/\partial\phi_i)(\partial\breve{\mu}_\ell/\partial\phi_j)}{v_\ell} \right\}$$

These first two steps can be proved using Lemma 5 and the results from Griffiths and Hajargasht (2015).

(iii) It can be shown that

(a) $$\lim_{N\to\infty} \sum_{\ell=1}^{N} \left\{ \frac{(\partial\kappa_\ell/\partial\phi_i)(\partial\kappa_\ell/\partial\phi_j)}{\kappa_\ell} \right\} \to \int \frac{(\partial f/\partial\phi_i)(\partial f/\partial\phi_j)}{f(x)^2} f(x)dx = E\left( \frac{\partial \ln f}{\partial\phi_i} \frac{\partial \ln f}{\partial\phi_j} \right)$$

(b) $$\lim_{N\to\infty} \sum_{\ell=1}^{N} \frac{(\partial\breve{\mu}_\ell/\partial\phi_i)(\partial\breve{\mu}_\ell/\partial\phi_j)}{v_\ell} \to 0$$

To prove (a), note that $\kappa_\ell = F(z_\ell;\boldsymbol{\phi}) - F(z_{\ell-1};\boldsymbol{\phi}) = \frac{\ell}{N} - \frac{\ell-1}{N} = \frac{1}{N}$. Thus, we can write:

$$\lim_{N\to\infty} \sum_{\ell=1}^{N} \left\{ \frac{(\partial\kappa_\ell/\partial\phi_i)(\partial\kappa_\ell/\partial\phi_j)}{\kappa_\ell} \right\} = \lim_{N\to\infty} \sum_{\ell=1}^{N} \frac{1}{N} \left\{ \frac{\partial F(z_\ell;\boldsymbol{\phi})/\partial\phi_i - \partial F(z_{\ell-1};\boldsymbol{\phi})/\partial\phi_i}{F(z_\ell;\boldsymbol{\phi}) - F(z_{\ell-1};\boldsymbol{\phi})} \frac{\partial F(z_\ell;\boldsymbol{\phi})/\partial\phi_j - \partial F(z_{\ell-1};\boldsymbol{\phi})/\partial\phi_j}{F(z_\ell;\boldsymbol{\phi}) - F(z_{\ell-1};\boldsymbol{\phi})} \right\}$$

Applying Taylor's theorem to both the numerator and the denominator of the following terms and assuming the necessary conditions on existence and boundedness of the derivatives of the density function $f$ we can show that

$$\frac{\partial F(z_\ell;\boldsymbol{\phi})/\partial\phi_i - \partial F(z_{\ell-1};\boldsymbol{\phi})/\partial\phi_i}{F(z_\ell;\boldsymbol{\phi}) - F(z_{\ell-1};\boldsymbol{\phi})} = \frac{\left(\partial f(z_{\ell-1};\boldsymbol{\phi})/\partial\phi_i\right) + O(1/N)}{f(z_{\ell-1};\boldsymbol{\phi}) + O(1/N)}$$

Assuming $f(z_{\ell-1},\phi) \neq 0$, some further manipulations lead to

$$\lim_{N\to\infty}\sum_{\ell=1}^{N}\left\{\frac{\left(\partial\kappa_\ell/\partial\phi_i\right)\left(\partial\kappa_\ell/\partial\phi_j\right)}{\kappa_\ell}\right\} = \lim_{N\to\infty}\frac{1}{N}\sum_{\ell=1}^{N}\left\{\frac{\partial f(z_{\ell-1},\boldsymbol{\phi})/\partial\phi_i}{f(z_{\ell-1},\boldsymbol{\phi})}\frac{\partial f(z_{\ell-1},\boldsymbol{\phi})/\partial\phi_i}{f(z_{\ell-1},\boldsymbol{\phi})}\right\}$$

Noting that $z_{\ell-1} = F^{-1}\left((\ell-1)/N\right)$, and using definition of a Riemann-Stieltjes integral, it can be seen that

$$\lim_{N\to\infty}\sum_{\ell=1}^{N}\left\{\frac{\left(\partial\kappa_\ell/\partial\phi_i\right)\left(\partial\kappa_\ell/\partial\phi_j\right)}{\kappa_\ell}\right\} = E\left\{\frac{\partial\ln f}{\partial\phi_i}\frac{\partial\ln f}{\partial\phi_j}\right\}$$

which proves (iii)(a). Using a similar argument, one can prove (iii)(b).

***Lemma 7:*** Consider the optimal GMM estimator $\hat{\boldsymbol{\phi}}$ in (8.1) from the moment conditions $\mathbf{H}(\boldsymbol{\phi}) = \tilde{\mathbf{z}} - \mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi})$. As $T\to\infty$, the variance of $\hat{\boldsymbol{\phi}}$ approaches the Cramer-Rao lower bound.

***Proof***: Using results in the proof of Proposition 8, and recognising that $c_i = F(z_i;\boldsymbol{\phi})$ we have

$$\operatorname{cov}\left(\hat{\boldsymbol{\phi}}\right) = \frac{1}{T}\left(\frac{\partial\mathbf{F}'^{-1}(\mathbf{c};\boldsymbol{\phi})}{\partial\boldsymbol{\phi}}\mathbf{W}_{L,11}\frac{\partial\mathbf{F}^{-1}(\mathbf{c};\boldsymbol{\phi})}{\partial\boldsymbol{\phi}'}\right)^{-1}$$

where

$$\mathbf{W}_{L,11} = D[\mathbf{f}(\mathbf{z};\boldsymbol{\phi})]\left[D\left(\mathbf{F}(\mathbf{z};\boldsymbol{\phi})\right) - \mathbf{F}(\mathbf{z};\boldsymbol{\phi})\mathbf{F}'(\mathbf{z};\boldsymbol{\phi})\right]^{-1}D[\mathbf{f}(\mathbf{z};\boldsymbol{\phi})]$$

From Lemma 2,

$$\frac{\partial\mathbf{F}'^{-1}(\mathbf{c};\boldsymbol{\phi})}{\partial\boldsymbol{\phi}} = -\frac{\partial\mathbf{F}'(\mathbf{z};\boldsymbol{\phi})}{\partial\boldsymbol{\phi}}D^{-1}[\mathbf{f}(\mathbf{z};\boldsymbol{\phi})]$$

And thus,

$$\operatorname{cov}\left(\hat{\boldsymbol{\phi}}\right) = \frac{1}{T}\left(\frac{\partial\mathbf{F}'(\mathbf{z};\boldsymbol{\phi})}{\partial\boldsymbol{\phi}}\left[D\left(\mathbf{F}(\mathbf{z};\boldsymbol{\phi})\right) - \mathbf{F}(\mathbf{z};\boldsymbol{\phi})\mathbf{F}'(\mathbf{z};\boldsymbol{\phi})\right]^{-1}\frac{\partial\mathbf{F}(\mathbf{z};\boldsymbol{\phi})}{\partial\boldsymbol{\phi}'}\right)^{-1}$$

Again using Lemma 5 it can be shown that

$$\mathrm{cov}(\hat{\boldsymbol{\phi}}) = \frac{1}{T}\boldsymbol{\Upsilon}^{-1}(\hat{\boldsymbol{\phi}}) \quad \text{where} \quad [\boldsymbol{\Upsilon}]_{ij} = \sum_{\ell=1}^{T}\left\{\frac{\left(\partial\kappa_\ell/\partial\phi_i\right)\left(\partial\kappa_\ell/\partial\phi_j\right)}{\kappa_\ell}\right\}$$

We showed in Lemma 6 that this quantity converges to the Cramer-Rao lower bound.

***Lemma 8:*** Consider the Lorenz curve estimation problem (8.4) with the moment conditions $\mathbf{H} = \left(\tilde{\mathbf{y}} - \mathbf{L}(\mathbf{c};\boldsymbol{\phi})\right)$ and individual data. As the number of observations approaches infinity, the variance of the optimal GMM estimator for $\boldsymbol{\phi}$ approaches the Cramer-Rao lower bound.

***Proof:*** According to Proposition 7, the asymptotic covariance for the optimal GMM estimator for this problem is equal to the asymptotic covariance of the optimal GMM estimator for the problem considered in Lemma 6. We showed in Lemma 6 that the latter asymptotic covariance converges to Cramer-Rao lower bound which proves the lemma.

**Proof of Proposition 9**: Lemma 7, Lemma 8 and the fact that asymptotic covariance of the maximum likelihood estimator is equal to Cramer-Rao lower bound proves the proposition.