Jenkins, Stephen P.; Burkhauser, Richard V.; Feng, Shuaizhang; Larrimore, Jeff

Working Paper

Measuring inequality using censored data: a multiple imputation approach

IZA Discussion Papers, No. 4011

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Jenkins, Stephen P.; Burkhauser, Richard V.; Feng, Shuaizhang; Larrimore, Jeff (2009) : Measuring inequality using censored data: a multiple imputation approach, IZA Discussion Papers, No. 4011, Institute for the Study of Labor (IZA), Bonn, https://nbn-resolving.de/urn:nbn:de:101:1-20090304434

This Version is available at:
https://hdl.handle.net/10419/35759

DISCUSSION PAPER SERIES

IZA DP No. 4011

Measuring Inequality Using Censored Data: A Multiple Imputation Approach

Stephen P. Jenkins
Richard V. Burkhauser
Shuaizhang Feng
Jeff Larrimore

February 2009

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Measuring Inequality Using Censored Data: A Multiple Imputation Approach

Stephen P. Jenkins
University of Essex and IZA

Richard V. Burkhauser
Cornell University

Shuaizhang Feng
Shanghai University of Finance and Economics, Princeton University and IZA

Jeff Larrimore
Cornell University

Discussion Paper No. 4011
February 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 4011
February 2009

ABSTRACT

Measuring Inequality Using Censored Data: A Multiple Imputation Approach*

To measure income inequality with right censored (topcoded) data, we propose multiple imputation for censored observations using draws from Generalized Beta of the Second Kind distributions to provide partially synthetic datasets analyzed using complete data methods. Estimation and inference uses Reiter's (*Survey Methodology* 2003) formulae. Using Current Population Survey (CPS) internal data, we find few statistically significant differences in income inequality for pairs of years between 1995 and 2004. We also show that using CPS public use data with cell mean imputations may lead to incorrect inferences about inequality differences. Multiply-imputed public use data provide an intermediate solution.

JEL Classification: D31, C46, C81

Keywords: income inequality, topcoding, partially synthetic data, CPS, Current Population Survey, Generalized Beta of the Second Kind distribution

Corresponding author:

Stephen P. Jenkins
Institute for Social and Economic Research
University of Essex
Wivenhoe Park
Colchester CO4 3SQ
United Kingdom
E-mail: stephenj@essex.ac.uk

* The research in this paper was conducted while Burkhauser, Feng and Larrimore were Special Sworn Status researchers of the U.S. Census Bureau at the New York Census Research Data Center at Cornell University. Conclusions expressed are those of the authors and do not necessarily reflect the views of the U.S. Census Bureau. This paper has been screened to ensure that no confidential data are disclosed. Support for this research from the National Science Foundation (award numbers SES-0427889 SES-0322902, and SES-0339191) and the National Institute for Disability and Rehabilitation Research (H133B040013 and H133B031111) is cordially acknowledged. Jenkins's research was supported by core funding from the University of Essex and the UK Economic and Social Research Council for the Research Centre on Micro-Social Change and the United Kingdom Longitudinal Studies Centre. We thank Ian Schmutte, Arnie Reznek, Laura Zayatz, the Cornell Census RDC Administrators, and all their U.S. Census Bureau colleagues who have helped with this project. Burkhauser thanks the Melbourne Institute for its hospitality while some of the research for this paper was undertaken. This paper has benefited from comments and suggestions made by John Micklewright, Steve Pudney and seminar participants at the Universities of Southampton and York.

1. INTRODUCTION

For assessing trends in the inequality of earnings or of household income inequality in the U.S.A., the March Current Population Survey (CPS) is the premier survey data source, widely used both within and outside government. (Administrative data are another source of information: see e.g. Piketty and Saez 2003 who use Internal Revenue Service tax data.) The CPS is, however, subject to an important limitation: the data are right censored ('topcoded'). To maximize confidentiality and to minimize disclosure risk, income values for each income source that are above a source-specific threshold are replaced in the public use data files by the threshold itself (the 'topcode'). Internal CPS data, used by the U.S. Census Bureau (various years) to produce official income distribution statistics, are also topcoded for the same reasons, albeit to a substantially lesser degree than the public use data. Right censoring is a problem for estimation of inequality levels because it suppresses genuine income dispersion, and it is a problem for estimation of inequality trends because CPS topcode values have not been adjusted consistently over time – the proportion of observations in the public use data with right censored values has fluctuated substantially over time. Topcoding also affects estimates of standard errors of inequality statistics because variance estimates depend on second- and higher-order moments, and their calculation is affected by right censoring. See Burkhauser, Feng, Jenkins and Larrimore (2008) for a recent review of topcoding practices in the March CPS and for references to earlier discussions of topcoding problems in CPS public use and internal data.

All previous imputation procedures applied to CPS data that we are aware of have used methods that yield a single imputation for each right censored value. And only rarely has the sampling variability of the estimates derived from the imputation-augmented data also been estimated in a manner that takes proper account of the right censoring. Instead, we propose a multiple imputation approach to estimating inequality using topcoded data following Reiter (2003). We show how this approach provides consistent estimates of not only inequality measures but also their sampling variances, accounting for both stochastic imputation error and sampling variability. We use the method to analyze recent trends in household income inequality in the U.S.A., exploiting our unprecedented access to internal CPS data.

Throughout the paper, income is defined in a conventional manner. It is pre-tax post-transfer household income excluding capital gains, adjusted for differences in household size using the square root of household size. (The specific procedures followed for constructing

the income measure are as discussed by Burkhauser and Larrimore, in press.) This income definition is common in the cross-national comparative income distribution literature (cf. Atkinson, Rainwater and Smeeding 1995) and studies of U.S. income distribution trends (cf. Gottschalk and Danziger 2005). Each individual is attributed with the size-adjusted income of the household to which he or she belongs. Income refers to income for the calendar year preceding the March interview. (All references to 'year' are to income year rather than survey year.) We convert the small number of negative and zero household income values each year to one dollar prior to our calculations because a number of inequality indices are defined only for positive income values. Our samples comprise all individuals in CPS respondent households, excluding individuals in group quarters or in households containing a member of the military. All statistics are calculated using the relevant CPS sampling weights. Sample sizes are large. For example, for 2004, the sample income distribution refers to 207,925 individuals in 75,660 households.

We compare our multiple imputation estimates of inequality levels and trends from the internal CPS data – what we label the Internal-MI series – with two series of estimates derived from public use CPS data. The Public-CM series arises when top-coded values are replaced by cell mean imputations derived from internal CPS data. These imputations have been provided by the U.S. Census Bureau for each year since 1995, and are available to all users of public use CPS data. The availability of this series is one reason why we restrict our attention to the period 1995–2004 in this paper. A second reason is that we wish to avoid any potential inconsistencies in the income series arising from the introduction of computer-assisted personal interviewing in the CPS in survey year 1994 (Ryscavage 1995). Third, it is well-known that U.S. income inequality increased substantially between the mid-1970s and the mid-1990s (see e.g. Danziger and Gottschalk 1995), and so we focus on a later period. For the decade starting from the mid-1990s, there is debate about the nature of inequality trends, but there is agreement that ascertaining trends in the very richest incomes is of particular importance in the U.S.A. and a number of other OECD countries (see e.g. Atkinson and Piketty 2007, Burkhauser, Feng, Jenkins and Larrimore 2008, and Piketty and Saez 2003).

The third series we analyze, labelled Public-MI, is also derived from public use CPS data, but applies a multiple imputation approach to estimation and inference that mimics the one that we apply to internal data. Although the internal data provide the best results, researchers can get access to them only under special conditions, whereas public use data are

available to all researchers. It is therefore of interest to explore the extent to which results from the Public-MI series match those from the Internal-MI one.

Using multiply-imputed internal CPS data, we show that the inequality of household income did not change significantly between 1995 and 2004, whether one uses ordinal evaluations based on Lorenz curves or cardinal comparisons based on a number of commonly-used inequality indices. We find that the cell mean augmented public use data lead to substantial under-estimates of inequality levels in every year, though the trends over time are tracked relatively well. However, the sampling variability of estimates derived from the cell-mean-augmented distributions is also under-estimated and, as a result, there is a tendency for inequality trends over the period to be shown (incorrectly) as statistically significant. Multiply-imputed public use data are shown to provide an intermediate case.

Although our research focuses on the case of right-censored data in the CPS, we would emphasize that the issues we have raised are applicable more widely – the CPS is not the only survey with topcoded data. For example, in the U.S.A., the National Longitudinal Survey of Youth topcodes some of its income sources as does the Panel Study of Income Dynamics. In the United Kingdom, in order to comply with the Statistics and Registration Services Act of 2007, the Annual Population Survey and the Quarterly Labour Force Survey have introduced topcodes on earnings data in their main public release files. In Germany, the wage data that are available from social insurance administrative registers are right censored at the earnings level corresponding to the upper limit to social insurance contributions.

2. RIGHT CENSORING IN INCOME DATA FROM THE MARCH CPS

In the March CPS, a respondent in each household is asked a series of questions on the sources of income for the household. Starting in 1975, respondents reported income from 11 sources, and since 1987 they have done so for income from 24 sources. High values for each separate income source are topcoded by the Census Bureau; it is not simply high total household income values that are topcoded. See Larrimore, Burkhauser, Feng and Zayatz (in press) for a full list of topcode values in the public use and internal CPS data, by income source.

An additional complication arises because household income is the aggregation of multiple income sources (across income types and household members), each of which may be topcoded. As a result, the prevalence of topcoding in total household income is

significantly greater than for any specific income source. For this reason, and also because income is measured using size-adjusted household income rather than nominal household income (see above), topcoded household income values are not necessarily the highest incomes – right censored observations may occur throughout the income distribution. Hence measuring inequality using the ratio of the 90th percentile to the 10th percentile (the 'P90/P10 ratio') with the goal of minimizing the impact of topcoding on inequality estimates will not be entirely successful: see Burkhauser, Feng and Jenkins (in press).

The proportion of individuals with topcoded household income in each March CPS from 1995 through 2004 is shown in Figure 1. In the public use data, the fraction is substantial, ranging between 2.1% and 5.7%. In the internal data, the proportion is roughly constant and small, only about .5%. The much lower prevalence of right censoring in the internal data indicates their substantial value for assessments of inequality compared to public use data. Using internal data rather than the public use data means that incomes are better measured for up to approximately 5.5% more observations. Nevertheless, censoring remains pervasive in the internal data. The mean size-adjusted household income value for topcoded observations in the internal data is around $200,000. The observation at the tenth percentile of the distribution of topcoded incomes is at the 55th percentile of the 1995 all-persons distribution, at the 87th percentile of the all-persons distribution for 2000, and somewhere between these ranks in the other years. So, accurate estimation of the degree of inequality needs to account for a non-trivial degree of right-censoring, even with CPS internal data. Our multiple imputation approach provides this.

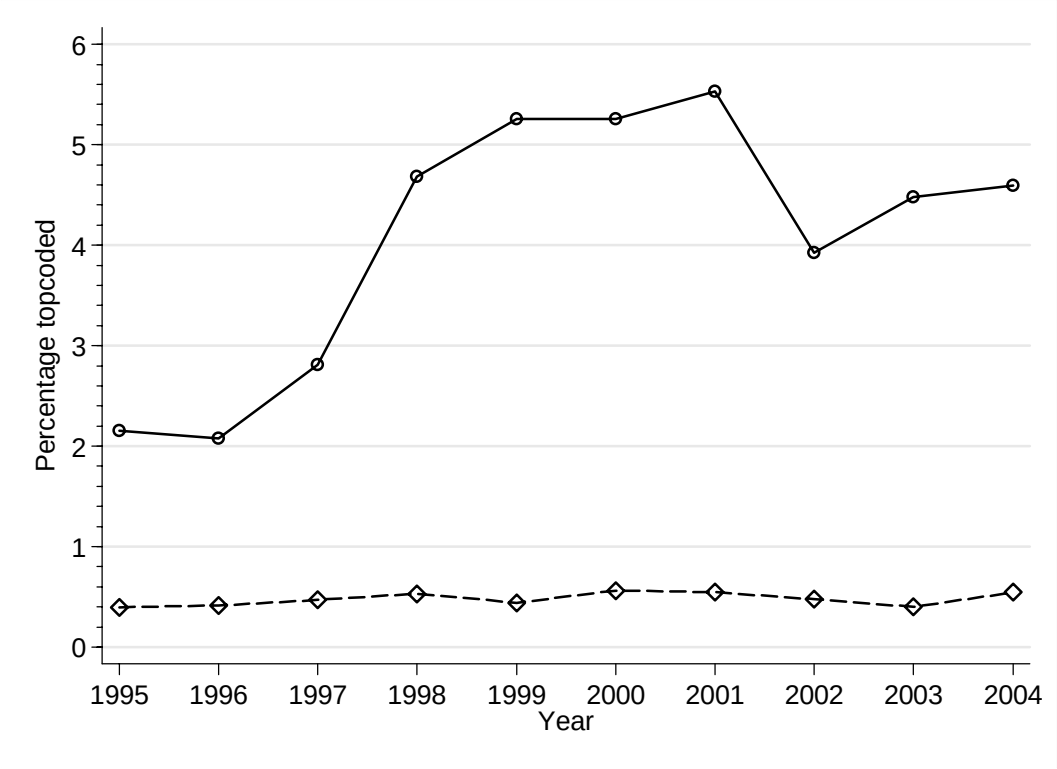


Figure 1. Percentage of individuals with topcoded household income in the March CPS public use data (solid line) and internal data (dashed line), by year. Authors' calculations from public use and internal March CPS data.

3. SINGLE IMPUTATION METHODS TO ACCOUNT FOR TOPCODING

The two principal imputation approaches to account for topcoding in public use CPS data that currently exist are reviewed in this section. The first approach is empirically based, using cell means derived from internal CPS data. (Another, more ad hoc, imputation procedure is to replace each topcoded value by a multiple of the topcode: see e.g. Katz and Murphy (1992), Lemieux (2006), and Autor, Katz and Kearney (2008).) The second approach is model based, assuming that the upper tail of the income distribution has a specific parametric functional form. These approaches yield a single augmented data set for analysis in which each topcoded value is replaced by a single imputed value: they are therefore 'single imputation' methods.

For each year of public use data since income year 1995 (survey year 1996), the Census Bureau has imputed a cell mean value to each topcoded value in the public use data. (Before 1995, the public use CPS data for each income source contained the topcode value for each income source for each observation topcoded on that source.) These imputations,

derived from internal CPS data, are, for each source, equal to the mean source income of all individuals with incomes greater than or equal to the topcode, subject to some constraints on minimum cell size. For labour income sources, the means are calculated within cells further divided by race, gender, and employment status. The Census Bureau initially provided cell means for wages and salaries, self-employment income and farm income, but later extended them to other non-governmental income sources (1998 and thereafter). Larrimore, Burkhauser, Feng and Zayatz (2008) provide further details of the derivation. They also distribute a consistent set of cell mean imputations to the wider research community that extends the Census Bureau series back to 1975.

These cell mean imputations are a substantial advance for analysis, providing more accurate measures of the incomes of topcoded observations than do the topcodes themselves. They have two limitations, however, which lead to underestimation of overall inequality statistics and their sampling variances. First, by construction, all observations within the same cell receive the same imputed value, thereby removing all within-cell income variation. Second, the cell means are derived from internal data which are themselves right censored. This imparts a downward bias to cell mean estimates of topcoded incomes (of unknown degree, since the actual incomes of the censored observations in the internal data are unknown), and hence also a downward bias in estimates of overall inequality and its sampling variance.

The second single imputation approach was developed before the Census Bureau cell mean series existed. Fichtenbaum and Shahidi (1988), using public use CPS data for 1967–1984, proposed that the upper tail of the U.S. income distribution for each year (specifically incomes greater than $100,000) be summarized by the one parameter Pareto distribution. The authors estimated the Pareto parameter for each year from grouped data published by the Census Bureau, and then used the properties of the Pareto distribution to calculate the mean income among the richest 5% and hence their share of total income, as well as the adjusted income shares of poorer income groups. Estimates of the Gini inequality index for each year were then derived from these income share data, and shown to be between .9% and 7.3% greater than corresponding Gini indices estimated ignoring topcoding. Essentially the same method was applied by Bishop, Chiou and Formby (1994) except that they used unit record public use CPS data, examined 1985–1989, and compared entire distributions using Lorenz curves as well as Gini coefficients. Notably, Bishop, Chiou, and Formby (1994) also estimated sampling variances for their inequality statistics, and used statistical inference procedures to test inequality differences. However, these procedures did not take account of

the additional variability introduced by the stochastic nature of the imputation process. (Also, neither Fichtenbaum and Shahidi (1988) nor Bishop, Chiou and Formby (1994) adjusted their family income measure to account for differences in family size and composition, and their estimation samples did not cover individuals from non-primary families.) A variation of the Pareto imputation method was also applied to public use CPS data on wages for 1979–1996 by Bernstein and Mishel (1997).

In addition to ignoring imputation uncertainty, the Pareto imputation method has the disadvantage, shared with the cell mean imputation approach, that only a single value is imputed to every topcoded observation in the relevant year, so income dispersion is under-estimated. Moreover, and as acknowledged by the authors cited, the goodness of the Pareto fit to CPS income data is debatable. On this, see also the critical discussion by Angle and Tolbert (1999). This suggests that the use of less restrictive parametric functional forms is productive in this context. Applications to public use CPS data on earnings include the two parameter gamma distribution (Angle 2003), the two parameter generalized Pareto (Stoppa) distribution (Burkhauser, Feng and Larrimore 2008), and the four parameter Generalized Beta of the Second Kind (GB2) distribution (Feng, Burkhauser and Butler 2006). In the next section, we make the case for applying the GB2 distribution to CPS data on household incomes, and for using a multiple rather than single imputation approach.

4. A MULTIPLE IMPUTATION APPROACH TO ACCOUNT FOR TOPCODING

Our multiple imputation approach consists of five steps, which we outline before discussing in more detail. First we fit an imputation model – a parametric functional form that is presumed to describe the income distribution in each year 1995–2004, including right censored observations. Second, for each observation with a right censored income, we draw a value from the distribution implied by the fitted model using an appropriate randomization procedure. Third, using the distribution comprising imputed incomes for censored observations and observed incomes for non-censored observations, we estimate our various inequality indices and associated sampling variances using complete data methods. Fourth, we repeat the second and third steps one hundred times for each year, and finally, we combine the estimates from each of the one hundred data sets for each year using the averaging rules proposed by Reiter (2003) for the case of 'partially synthetic' data. This accounts for the uncertainty added to estimates by the imputation process as well as for sampling variability. Application of the five-step approach to internal CPS data yields our

Internal-MI series of estimates; application to public use CPS data yields our Public-MI series.

Ours is the only study that we are aware of that has applied multiple imputation methods to right censored data for the purposes of analyzing income inequality. The closest study to ours is An and Little (2007). They fit lognormal and power-transformed normal distributions to data from the 1995 Chinese household income project, and use the estimates to multiply impute incomes to topcoded observations. Their focus was on estimation of and inference about mean incomes and income regressions for a single year rather than estimates of income inequality and trends. Gartner and Rässler (2005) used lognormal distributions fitted to German wage data for 1991–2001 to multiply impute values for topcoded observations. Again the focus was estimation and inference concerning mean incomes and income regressions rather than income inequality.

We assume that the distribution of size-adjusted household income in each year is described by the four parameter Generalised Beta of the Second Kind (GB2) distribution (McDonald 1984), with probability density function

$$f(y) = \frac{a y^{ap-1}}{b^{ap} B(p,q)\left[1+(y/b)^{a}\right]^{p+q}}, y > 0$$

and cumulative density function (CDF)

$$F(y) = I(\, p, q, (y/b)^{a} / [1 + (y/b)^{a}] \,), y > 0$$

where parameters a, b, p, q, are each positive. $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p + q)$ is the Beta function, $\Gamma(.)$ is the Gamma function, and $I(p, q, x)$ is the regularized incomplete beta function also known as the incomplete beta ratio. Parameter b is a scale parameter, and a, p, and q are each shape parameters. The GB2 is a flexible functional form incorporating many distributions as special cases. For example, the Singh-Maddala (Burr type 12) distribution is the special case of the GB2 distribution when $p = 1$; the Dagum (Burr type 3) distribution is the special case when $q = 1$; and the lognormal distribution is a limiting case. For details, see McDonald (1984) and Kleiber and Kotz (2003). Many studies have shown that the GB2 model fits income distributions extremely well across different times and countries: see inter alia McDonald (1984), Bordley, McDonald and Mantrala (1996), Brachmann, Stich and Trede (1996), Bandourian, McDonald and Turley (2003), and Jenkins (in press).

Of particular importance in the current context is the desirable behaviour of the GB2 distribution in its upper tail. Consistent with extreme value theory, the upper and lower tails lie in the domain of attraction of the Fréchet distribution. The upper tail is regularly varying

(with variation parameter equal to $-aq$) and it is heavy in that it decays like a power function as income increases, rather than decaying exponentially fast (as for the log-normal distribution, with middle heavy upper tail), or polynomial decreasing (as for Pareto distributions). See Schluter and Trede (2002, Appendix A) and Kleiber and Kotz (2003) on regular variation concepts and the upper tail behaviour of GB2 and other distributions.

We estimate the GB2 distribution parameters by maximum likelihood (ML), separately for each year 1995–2004. To ensure that model fit was maximized at the top of the distribution, we fit each GB2 distribution using observations in the richest 70 percent of the distribution only, making appropriate corrections for left truncation in the ML procedure. (We chose the 30th percentile as the left truncation point after experiments balancing goodness of fit with ease of maximization.) We specify the sample log-likelihood for each year's data as

$$\ln L = \sum_{i=1}^{N} w_i \left\{ \frac{c_i \ln[1 - F(y_i)] + (1 - c_i) \ln[f(y_i)]}{1 - F(z)} \right\}$$

where $i = 1,\ldots, N$, indexes each individual sample observation, w_i is the sample weight for i, and $c_i = 1$ if i is an observation with a right censored household income value, and $c_i = 0$ otherwise. The denominator of the expression adjusts for left truncation: z is the income level corresponding to the left truncation point. For maximization, we use the modified Newton-Raphson procedure implemented in Stata's *ml* command (StataCorp, 2005), with the parameter covariance matrix estimates based on the negative inverse Hessian. Convergence was achieved easily within several iterations. For brevity, we do not report estimates for each year but they are available from the authors on request.

For the internal data, model fit varied slightly across years, but was generally excellent. This is demonstrated first by the precision of the parameter estimates. For example, the smallest *t*-ratio for any parameter estimate (always for p) was greater than seven, and was typically at least two or three times larger for a and for q. Wald tests of parameter values suggested that we could easily reject restricted models such as the Singh-Maddala or Dagum distributions in favour of the GB2 distribution. Excellent goodness of fit to the internal data is also demonstrated by the probability plots shown in Figure 2 for each year. These are plots of the cumulative probabilities of income expected given the estimated GB2 parameters against the cumulative probabilities of income observed in the data. (Each chart is based on the richest 70% of each distribution for each year, for the reasons explained earlier.) Excellent goodness of fit is demonstrated by the fact that every plot lies extremely close to a 45° ray

from the origin. Although fitted values must lie above observed values over the probabilities corresponding to right censored observations in the internal data (at the very far right of each chart), we note that there is also no perceptible change in the nature of the plots for probabilities in the neighbourhood of these observations. Such smooth continuity increases our confidence in the use of the GB2 for imputing incomes to right censored observations in the internal CPS data.

We also fit GB2 models separately to public-use data for each year, using procedures that mimicked those for the internal data. Given the greater prevalence of right-censoring, model goodness-of-fit was not quite as good as for the internal data, but very good nonetheless. This is illustrated by the probability plots shown in Appendix Figure A.

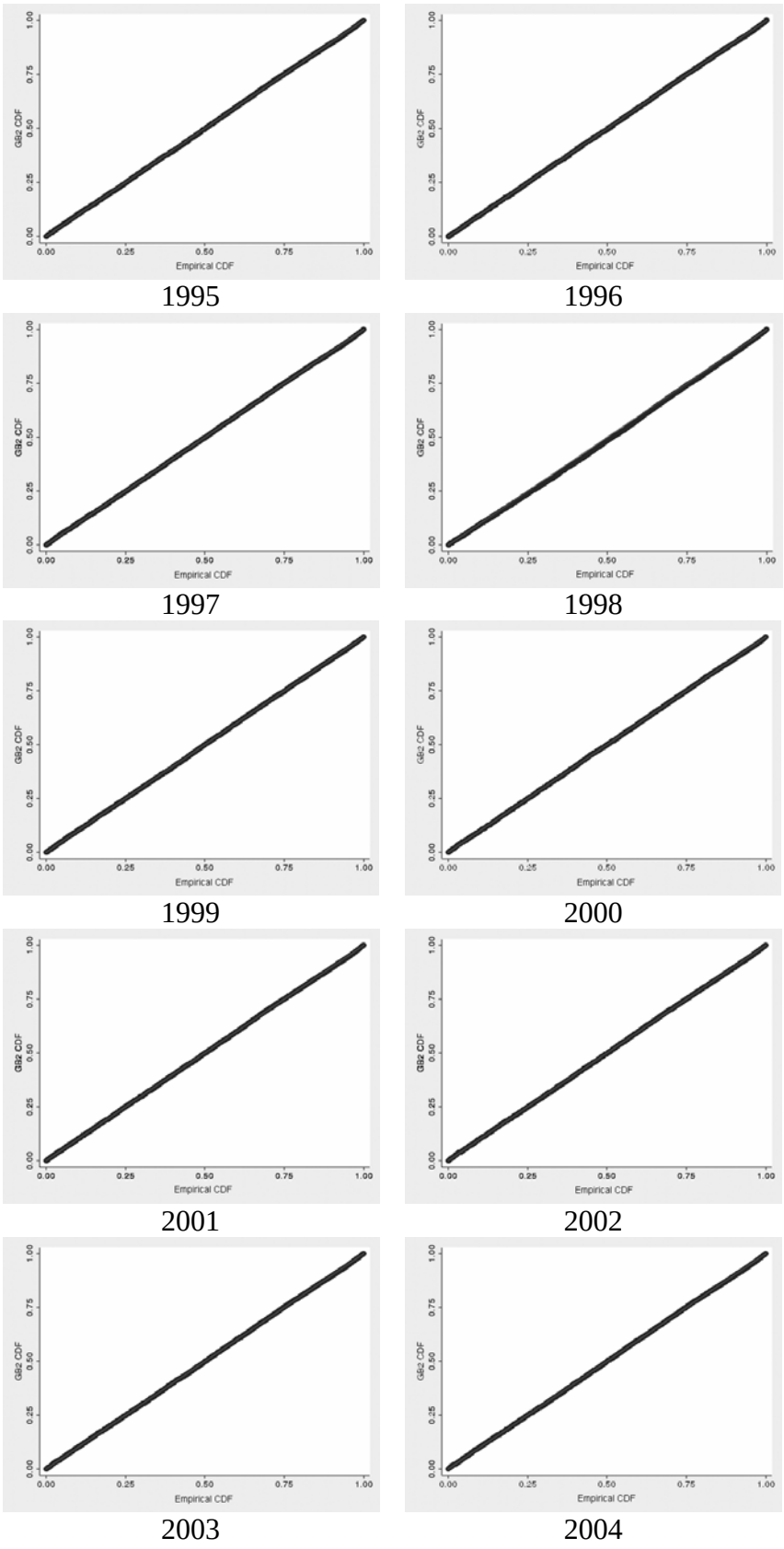


Figure 2. Probability plots for GB2 estimates: fitted versus observed. Plot based on richest 70% of each distribution only. GB2 estimates account for left-truncation and right-censoring (see text). Source: authors' derivations from internal CPS data.

The second stage of our multiple imputation approach uses the GB2 estimates to derive imputed values for topcoded observations for each year using the inverse transform sampling method. Given fitted GB2 CDF, $\hat{F}(y)$, the corresponding CDF for topcoded observation *i* is, using standard formulae for truncated distributions:

$$\hat{G}(y_i) = [\hat{F}(y_i) - \hat{F}(t_i)]/[1 - \hat{F}(t_i)]$$

where t_i is the topcode for *i*, and y_i is the 'true' value for that observation (which we are unable to observe). Letting $u_i = \hat{G}(y_i)$, and inverting the expression for the income distribution among topcoded observations, we have

$$y_i = \hat{F}^{-1}\left(u_i[1 - \hat{F}(t_i)] + \hat{F}(t_i)\right).$$

A value of y_i for each topcoded observation is generated by substituting into this expression a value of u_i equal to a random draw from a standard uniform distribution. The combination of the observed incomes for non-topcoded observations with the imputed incomes for topcoded observations produces a partially synthetic data set for each year to which we can apply complete data methods to estimate our inequality statistics of interest. (Fully synthetic data consist of entirely multiply imputed data.) Repetition of the process $m > 1$ times produces *m* partially synthetic data sets for each year and, correspondingly, *m* sets of inequality estimates for each year which we combine in a manner discussed shortly. Note that the observations without censored data are common across each of the *m* partially synthetic data sets. Clear cut rules for the choice of *m* do not exist, but the number used is often relatively small (10 or fewer). In the next section, we report estimates based on $m = 100$. (In preliminary research, we used $m = 20$ and derived similar conclusions to those reported here.)

For inference from our multiply-imputed partially synthetic data sets, we use the combination formulae derived by Reiter (2003), as follows. Suppose that inference is required about some scalar *Q*, where *Q* is a measure of inequality such as the Gini index, and index the partially synthetic data sets by $j = 1, \ldots, m$. Denote the point estimator of *Q* from partially synthetic data set *j* by q_j and the estimator of its variance by v_j. Reiter (2003) shows that one should estimate *Q* using the mean of the point estimators

$$\bar{q}_m = \sum_{j=1}^{m} q_j / m$$

and use

$$T_p = (b_m / m) + \bar{v}_m$$

to estimate the variance of $\bar{q}_m$, where

$$b_m = \sum_{i=1}^{m} (q_i - \bar{q}_m)^2 / (m-1) \quad \text{and} \quad \bar{v}_m = \sum_{j=1}^{m} v_j / m .$$

Thus, the multiple imputation point estimate is the simple average of the point estimates derived using complete data methods from each of the *m* partially synthetic data sets. The variance of this estimate is the average of the sampling variances plus a term reflecting the finite number of imputations, *m*. T_p differs from Rubin's (1987) rule for the combination of estimates in the fully synthetic data case, in which case $T_p = b_m + b_m / m + \bar{v}_m$. The expression for the fully synthetic data case includes additional variability (the term b_m) to average over the response mechanism (Rubin 1987). By contrast, '[t]his additional averaging is unnecessary in partially synthetic data settings since the selection mechanism … is not treated as stochastic' (Reiter 2003, p. 5). The selection mechanism in our case refers to the choice of topcodes by the Census Bureau for the CPS. For large sample sizes, inference concerning *Q* can be based on *t*-distributions with degrees of freedom $v_p = (m-1)(1 + r_m^{-1})^2$ where $r_m = (m^{-1} b_m / \bar{v}_m)$. Because our sample sizes are large and *m* is large, v_p is very large also, so the *t*-distribution is approximated very well by a normal distribution and that is what we use for inference.

Because cardinal indices of inequality differ in their sensitivities to income differences in different ranges of the income distribution (Atkinson 1970), we estimate inequality indices that reflected this feature in a systematic way. Specifically, we consider the mean log deviation, Theil index, and half the coefficient of variation squared, plus the Gini coefficient. The first three indices belong to the one parameter Generalized Entropy class GE(α) with parameter α = 0, 1, 2, respectively, and range from being bottom-sensitive (MLD) to being sensitive to income differences at the top of the distribution ($CV^2/2$). The commonly-used Gini coefficient is a middle-sensitive inequality index. (For a review of inequality index properties and index formulae, see Cowell 2000.) We computed distribution-free variance estimates for the inequality indices according to formulae provided by Biewen and Jenkins (2006) for GE indices and by Kovačević and Binder (1997) for the Gini index. In both cases, we account for the clustering of individuals in households and for the stochastic sample weights. The computations were undertaken using the Stata modules provided by Biewen and Jenkins (2005) and Jenkins (2006).

We also checked whether our estimates of inequality trends were robust to the choice of inequality measure by employing Lorenz dominance analysis, checking whether or not Lorenz curves of income distributions for pairs of years crossed. In this case, the statistics of

relevance for each year are the cumulative shares of income for different income groups for the sample ranked in ascending order of income, and their sampling variances. If there is statistically significant Lorenz dominance, then there is a unanimous ordering of income distributions according to all standard inequality indices (Foster 1985). These include the four indices mentioned in the previous paragraph. Again we account for the clustering of individuals in households and for the stochastic sample weights. The computations of the distribution-free variance formulae provided by Kovačević and Binder (1997) were undertaken using the Stata module provided by Jenkins (2006). Following common practice, the income shares were computed at the 19 vingtiles.

Reiter (2003) provides a Bayesian derivation for his data combination inference formulae, and also the two conditions under which the inferences are valid from a frequentist perspective: that the analyst uses randomization valid estimators and that the synthetic data generation methods are proper in a sense similar to Rubin (1987). Of necessity we argue that these conditions are satisfied in our case, though note that it is impossible to test their validity since access to the actual values for topcoded observations in the internal CPS data is impossible. (Reiter developed his combination rules under the assumption that, in effect, the producer of the multiply imputed partially synthetic data sets had used imputation methods that satisfy the conditions.) Our imputation procedures are randomization based though not fully Bayesian since we did not draw from the posterior predictive distribution of the GB2 parameters in each year. Because of the complexity of implementing this procedure in our context, we drew from the full data posterior distribution, treating the GB2 parameters as known and appeal to the excellent fit of our GB2 models. An and Little (2007) employed the same procedure when they derived multiple imputations for data assumed to follow a power-transformed normal distribution.

5. ESTIMATES OF U.S. INCOME INEQUALITY, 1995–2004

We first discuss the results from our analysis of Lorenz dominance, and then the inequality indices. Throughout, tests for statistically significant differences are based on changes between distributions at least two years apart. We do not test for differences between adjacent years because of the rotation group structure of the CPS (about half of the sample is the same in the March CPS for consecutive years): our test procedures are predicated on having independent samples in each year. The focus is on the Internal-MI series, but supplemented by discussion of how the estimates from the Public-CM and Public-MI series

compare. Although the latter two series under-estimate inequality measures and associated sampling variances (as explained earlier), it is of interest to know whether these features lead to erroneous conclusions about inequality differences.

Detailed results from the Lorenz dominance analysis of the Internal-MI data for the beginning, middle and end of the period (1995, 2000, and 2004) are reported in Table 1. The details for the other years are not shown for brevity. Shown are the estimated Lorenz ordinates (cumulative income shares) at each successive twentieth of the distributions, together with the estimated standard errors (in parentheses) derived using the methods discussed in the previous section. The rightmost three columns report distribution-free Lorenz dominance test statistics for pairwise comparisons between the three years. For a pairwise comparison between year *A* and year *B*, and income group $k = 1, \ldots, 19$, each test statistic (Δ_k) is

$$\Delta_k = \left(\hat{L}_k^A - \hat{L}_k^B\right) \Big/ \sqrt{\hat{V}_k^A + \hat{V}_k^B} \ ,$$

where $\hat{L}_k^A$ is the estimate of the k^{th} Lorenz ordinate, $\hat{V}_k^A$ is the estimate of its variance, and correspondingly for year *B*. Hypothesis testing uses the multiple comparison union-intersection method of Bishop, Formby and Smith (1991*a*, 1991*b*) and Bishop, Chiou and Formby (1994). Tests are based on a 5% significance level and take account of the fact that each dominance test is based on 19 simultaneous tests. The critical value is therefore obtained from the Student maximum modulus distribution (Beach and Richmond 1985): SMM(19,∞) = 3.01.

There are four possible outcomes from each set of tests associated with the comparison of years *A* and *B*, as Bishop and colleagues explain. First, there may be no statistically significant difference between any pair of Lorenz ordinates, in which case *A* and *B* are ranked as equivalent in terms of inequality (i.e. equality is taken as the null hypothesis): $|\Delta_k| \leq 3.01$ for all *k*. Second, if there are positive and statistically significant differences in ordinates and no negative and statistically significant differences, then *A* Lorenz dominates *B*: inequality is lower according to all standard inequality indices ($\Delta_k > 3.01$ for some *k* and $|\Delta_l| \leq 3.01$ for $l \neq k$). The reverse is the case, third, if there are negative and statistically significant differences in ordinates and no positive and statistically significant differences ($\Delta_k < -3.01$ for some *k* and $|\Delta_l| \leq 3.01$ for $l \neq k$). Fourth, if there are negative and positive differences that are statistically significant, the Lorenz curves cross and a unanimous inequality ranking cannot be derived ($\Delta_k > 3.01$ for some *k* and $\Delta_l < -3.01$ for some $l \neq k$).

Table 1. Lorenz Ordinates, Standard Errors, and Test Statistics for Pairwise Lorenz Comparisons (Internal-MI series)

Cumulative population share	Year			SMM test statistics (Δ_k)		
	1995	2000	2004	2004 vs. 1995	2004 vs. 2000	2000 vs. 1995
.05	.0040	.0040	.0031	–.1167	–.1018	–.0006
	(.0060)	(.0074)	(.0053)			
.10	.0134	.0137	.0122	–.1587	–.1636	.0234
	(.0060)	(.0074)	(.0053)			
.15	.0266	.0268	.0250	–.2084	–.2042	.0206
	(.0059)	(.0074)	(.0053)			
.20	.0433	.0432	.0411	–.2746	–.2378	–.0025
	(.0059)	(.0073)	(.0053)			
.25	.0633	.0629	.0603	–.3722	–.2870	–.0369
	(.0059)	(.0073)	(.0052)			
.30	.0866	.0857	.0827	–.4988	–.3427	–.0891
	(.0058)	(.0073)	(.0052)			
.35	.1132	.1118	.1083	–.6344	–.3958	–.1510
	(.0058)	(.0072)	(.0052)			
.40	.1431	.1407	.1372	–.7760	–.3958	–.2684
	(.0057)	(.0072)	(.0051)			
.45	.1764	.1731	.1694	–.9233	–.4153	–.3717
	(.0057)	(.0071)	(.0051)			
.50	.2133	.2089	.2052	–1.0692	–.4207	–.4867
	(.0056)	(.0071)	(.0050)			
.55	.2538	.2483	.2447	–1.2269	–.4180	–.6188
	(.0055)	(.0070)	(.0050)			
.60	.2982	.2915	.2881	–1.3729	–.3981	–.7575
	(.0055)	(.0069)	(.0049)			
.65	.3467	.3389	.3360	–1.4774	–.3456	–.8917
	(.0054)	(.0069)	(.0048)			
.70	.4000	.3908	.3888	–1.5692	–.2400	–1.0665
	(.0053)	(.0068)	(.0048)			
.75	.4587	.4480	.4476	–1.5676	–.0432	–1.2518
	(.0052)	(.0067)	(.0047)			
.80	.5240	.5116	.5122	–1.7107	.0653	–1.4705
	(.0051)	(.0066)	(.0046)			
.85	.5975	.5835	.5859	–1.7197	.2967	–1.6982
	(.0050)	(.0065)	(.0045)			
.90	.6824	.6674	.6715	–1.6448	.5387	–1.8687
	(.0049)	(.0064)	(.0044)			
.95	.7862	.7713	.7769	–1.4738	.7473	–1.9292
	(.0046)	(.0062)	(.0043)			

NOTE: Standard errors are in parentheses. Source: authors' calculations from internal CPS data.

Table 1 suggests that, according to the point estimates of the ordinates, the Lorenz curve moved slightly outwards between 1995 and 2004 (indicating greater inequality). However, all the test statistics Δ_k are smaller than 3.01, and so we cannot reject the null hypothesis of equality of ordinates. For the two subperiods, the pattern of change in the ordinates is more complex – there are both positive and negative changes in the ordinates – but the outcome of the pairwise dominance test is the same. Indeed, we cannot reject the null hypothesis of no statistically significant difference between Lorenz ordinates for all 36 pairwise comparisons undertaken (using all pairs of years 1995–2004 excluding adjacent years). Thus, according to Lorenz dominance tests applied to multiply imputed internal data, there was no significant change in inequality within and over the period 1995–2004 according to all standard inequality indices.

We repeated the Lorenz dominance tests using the Public-CM data and found the same result, with one difference. That is, we could not reject the null hypothesis of no statistically significant differences between Lorenz ordinates for all pairwise comparisons undertaken, with the exception of comparisons involving 1999 which was apparently more equal than any of the other years considered. For example, in the comparison between 1999 and 2004, $\Delta_k > 3.01$ for $k = 15, 16, 17, 18, 19$, and $0 < \Delta_k < 3.01$ otherwise. That is, cumulative income shares were significantly lower in the top quarter of the income distribution in 2004 compared to 1999. Apart from the results for 1999, there is consistency between the conclusions derived from the Internal-MI and Public-CM series.

Lorenz dominance tests based on the Public-MI data led to slightly different results. Again, there was less inequality in 1999 than in 1995 and in 1996 and 2003. In addition, 1995 was more equal than 2003. There were also a number of Lorenz curve crossings (case 4 above). Three of these involved 1999 (with 2001, 2003 and 2004); the fourth involved 1995 and 2000. Aside from the results from 1999, there is broad consistency between these estimates and those from the Internal-MI series.

Estimates of inequality indices and test statistics for pairwise comparisons for selected years are shown in Table 2 for the Internal-MI series. The test statistics are for pairwise difference-in-means *t*-tests, and so the relevant critical value using a 5% significance level is approximately 1.96. We find that the estimate of each index increased between 1995 and 2000 and between 2000 and 2004. The estimated increase between 1995 and 2004 is largest for the GE(2) and GE(0) indices (20% and 18%, respectively), and smallest for the Gini and GE(1) indices (3% and 9%, respectively). However it is only for the Gini and GE(0) indices

that the increases are statistically different from zero. (Finding significant differences for a specific inequality index is consistent with the Lorenz dominance test results reported above, because the latter were only based on comparisons at 19 income values rather than all sample values.) Few subperiod increases are statistically significant either – the exceptions mainly concern GE(0).

Table 2. Inequality Indices, Standard Errors, and Test Statistics for Pairwise Comparisons (Internal-MI series)

Index	Year			Test statistics		
	1995	2000	2004	2004 vs. 1995	2004 vs. 2000	2000 vs. 1995
Gini	.4312	.4426	.4450	3.1020	.4569	2.1027
	(.0033)	(.0043)	(.0030)			
GE(0)	.4017	.4296	.4751	7.6965	4.1924	2.6074
	(.0066)	(.0084)	(.0069)			
GE(1)	.3732	.4156	.4051	1.7503	–.4388	1.7852
	(.0127)	(.0200)	(.0130)			
GE(2)	.9658	1.4101	1.1600	.5650	–.4235	.8079
	(.1894)	(.5163)	(.2867)			

NOTE: Standard errors are in parentheses. Source: authors' calculations from internal CPS data.

The complete set of tests for pairwise inequality index differences are summarized in Table 3 for all three series. For each year, inequality index, and data source series, the cell entry shows the year(s) for which there is a statistically significant difference in inequality between the comparison year (column 1) and the year(s) shown. A blank cell means no comparison between that year and any other year is statistically significant. The table points to several findings about inequality differences and about consistency in results across series.

First, according to the gold standard of the Internal-MI series, few inequality differences are significantly different from zero. Where they are statistically significant, they typically refer to differences between the beginning of the period and the end of the period. The greatest number of statistically significant differences refer to the GE(0) index. Also, none of the tests for differences in the top-sensitive index GE(2) based on the Internal-MI series are statistically significant.

Table 3. Pairwise Comparisons of Inequality Indices, by Index, Year and Estimate Series

Year	Gini			GE(0)			GE(1)			GE(2)		
	Internal-MI	Public-CM	Public-MI	Internal-MI	Public-CM	Public-MI	Internal-MI	Public-CM	Public-MI	Internal-MI	Public-CM	Public-MI
1997		1995	1995	1995	1995	1995		1995	1995		1995	
1998			1995, 1996	1995,1996	1995,1996	1995, 1996			1995, 1996			
1999		1995–1997	1995–1997		1997	1995, 1996		1995–1997	1995–1997		1995–1997	
2000	1995		1995, 1996	1995	1995	1995, 1996			1995–1996		1997	
2001	1995, 1996, 1999	1995, 1996, 1998, 1999	1995, 1996, 1999	1995–1999	1995–1999	1995, 1996	1995, 1999	1995, 1998, 1999	1995, 1999		1999	
2002	1995	1995, 1999	1995, 1996, 1999	1995, 1996, 1999	1995, 1996, 1999, 2000	1995, 1996, 2000		1995, 1999	1995, 1996, 1999		1995, 1999, 2000	
2003	1995	1995, 1998, 1999	1995–1997, 2001	1995–2000	1995–2000	1995–2001	2001	1995, 1999	1995, 1996		1999	
2004	1995	1995, 1996, 1998–2000	1995, 1996, 1999	1995–2000, 2002	1995–2002	1995–2002		1995, 1998, 1999, 2000	1995, 1996, 1999		1995, 1998–2000	

NOTE: For each year, inequality index, and data source series, the cell entry shows the year(s) for which there was a statistically significant difference in inequality between that year and the year(s) shown. A blank cell means no comparison between that year and any other year was statistically significant. Comparisons undertaken for every pair of years 1995–2004, adjacent years excepted. Authors' calculations from internal and public use CPS data.

Second, we find more statistically significant differences using the Public-CM series than the Internal-MI series. For some reference years when comparisons based on the Internal-MI series yield no statistically significant differences at all, there are statistically significant differences according to the Public-CM series. Consider, for example, the comparisons for each of the reference years 1996–1999 for the Gini and GE(1) indices, and all comparisons for GE(2). And, whenever there is any statistically significant difference concerning a year *A* and a year *B* according to the Internal-MI series, there are often statistically significant differences between year *A* and additional years as well according to the other two series. For example, according to the Internal-MI series, there was a statistically significant difference between the Gini indices for 2004 and 1995. According to the Public-CM series, the Gini index for 2004 differed from the estimates for 1995, 1996, 1998, 1999 and 2000. In only two instances was there a statistically significant difference according to the Internal-MI series but not the Public-CM one: the Gini comparison for 2000 and 1995, and the GE(1) comparison for 2003 and 2001.

The explanation for these findings is that the suppression of genuine within-cell income dispersion that is associated with cell mean imputations lead to underestimation of not only inequality indices but also their sampling variances. There is therefore a tendency for estimated inequality trends based on the Public-CM series to be judged statistically significant when they are not.

Like the Public-CM series, the Public-MI series of estimates generally leads to more statistically significant differences than does the Internal-MI one for any given reference year. Observe, for example, that according to the Public-MI series, inequality in 1995 differs from inequality in every other year compared, according to all indices except GE(2). There are, however, some years and indices for which the Public-MI series shows a significant difference and the Internal-MI one does not. See, for instance, the results for reference years 2001–2003 and GE(0), and reference year 2001 and GE(1). There is also not complete consistency between the patterns of pairwise differences found for the Public-MI and Public-CM series. This is unsurprising, given the different ways in which the estimates were derived.

The differences between the three series are highlighted by Figure 3. This shows the estimates for each of the four inequality indices, year by year, together with the associated 95% confidence intervals. Unsurprisingly, estimates of inequality levels in each year are greater for the Internal-MI series than for the Public-CM series. However, both series point to similar trends over the period: they suggest that inequality levels fell slightly at the end of the 1990s, especially between 1998 and 1999, and again between 2001 and 2002. This

consistency appears reassuring for analysts, especially since all researchers have access to cell mean-augmented public use CPS data whereas access to internal CPS data is subject to special conditions.

However, Figure 3 also clearly shows that confidence intervals for inequality indices estimated using the Public-CM data are too narrow by a substantial amount. This feature is particularly striking for the top-sensitive GE(2) index, which is not surprising because it is at the very top of the distribution that the data series differ in dispersion. Part of the greater fluctuation in inequality levels and wider confidence intervals in the Internal-MI series may reflect the GE(2)'s relatively greater non-robustness to the effects of outliers in the sense discussed by Cowell and Victoria-Feser (1996). However, we would argue that the patterns shown in Figure 3 primarily reflect sampling and imputation variability, since the averaging process used to combine the estimates from our 100 multiply imputed data sets are likely to smooth out the effects of any outliers being added by the imputation process. In sum, the reassuring consistency between the Internal-MI and Public-CM series evaporates if the researcher is interested in statistical inference and not simply point estimates.

The Public-MI estimate for any given year and index lies between the corresponding Internal-MI and Public-CM estimate, and is generally closer to the latter rather than the former. (There are a few exceptional cases in which the Public-MI estimate is slightly smaller than the Public-CM estimate, most of which involve the estimates for 1995 or 1996.) Our explanation for lower inequality in the Public-MI series than the Internal-MI one is that the imputation model underlying the Public-MI series does not work as well as the model underlying the Internal-MI series and this, in turn, is related to the substantially greater prevalence of topcoded data in the public use data compared to the internal data. It is the same feature that leads to confidence intervals that are smaller than for Internal-MI series. (They are, however, larger than for the Public-CM series as the construction of the Public-MI series incorporates more variability via its imputation process.) Our explanation for the Public-MI estimates being relatively close to the Public-CM one is that derivation of the Public-MI series did not use any information from the internal data, whereas the Public-CM series did. The impetus to inequality that is added by the randomization process in the derivation of the former source is matched by the use of information about actual incomes from the internal data in the latter source.

The final feature of Figure 3 that we wish to comment on is the results for 1999. Relative to trend, this appears to be an outlier year, and echoes results from the Lorenz dominance analysis reported earlier. Interestingly, the index estimates for 1999 are above

those for immediately previous and succeeding years according to the Public-MI series, but below them according to the other two series. Indeed, if the results for 1999 were discarded, the trends for the three series would look much more similar. We do not have a complete explanation for the exceptional 1999 results. We rule out differences in the imputation process for the public data for that year, because we can see no clear differences between the parameter estimates of the GB model for 1999 and those for the years before and afterwards. And the computer code used to implement the imputation randomization process is generic. Since the Internal-MI and Public-CM series both rely on internal data for their derivation, whereas the Public-MI series does not, we suspect that there is some feature of the internal data for that year that underlies the pattern.

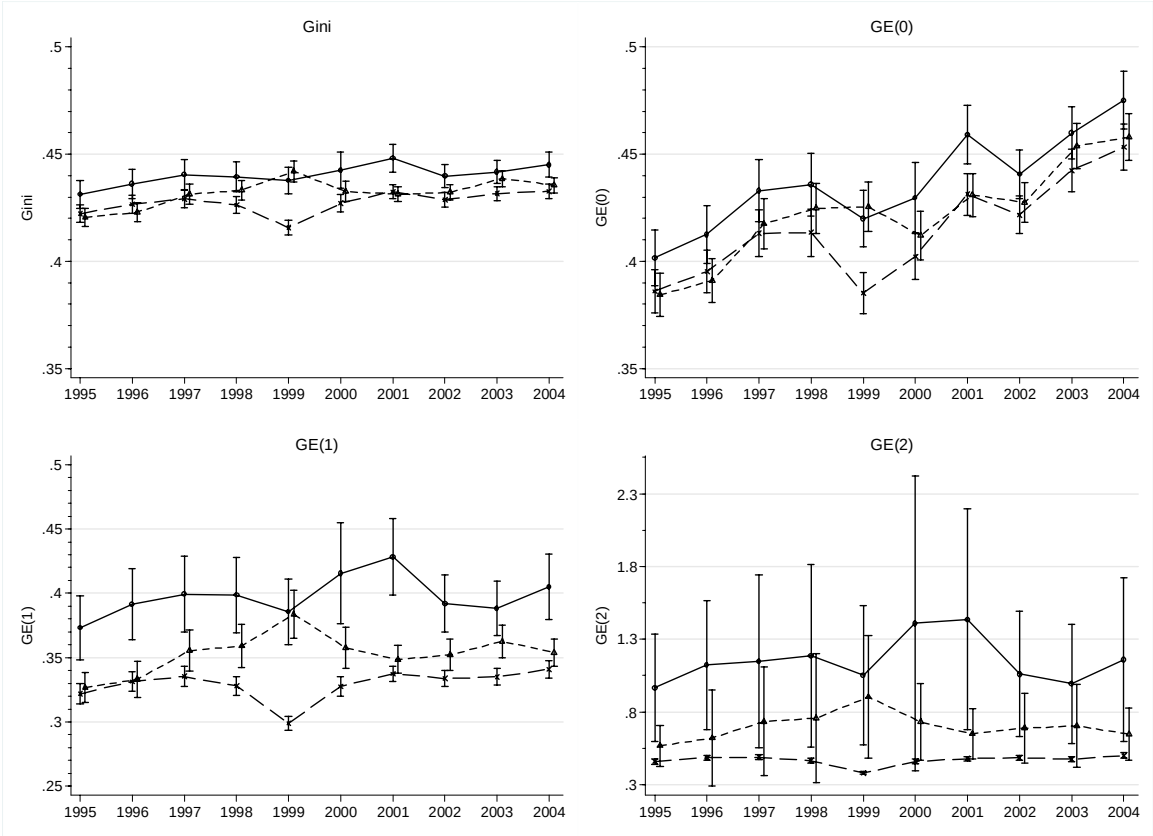


Figure 3. Inequality indices with 95% confidence intervals, 1995–2004. Internal-MI series (solid line), Public-CM series (long-dashed line), and Public-MI series (short-dashed line), derived from internal and public use CPS data.

6. DISCUSSION

We have demonstrated how a multiple imputation approach may be used to estimate inequality levels and trends from right censored income data. With a suitable imputation model, researchers may impute values to topcoded observations, thereby creating multiple partially synthetic data sets to be analyzed using complete data methods. Estimates combined using straightforward formulae can be used for statistical inference.

Applying the multiple imputation approach to internal data from the CPS, we have shown that no clear cut conclusions about the changes in income inequality over the period between 1995 and 2004 can be drawn. According to Lorenz dominance analysis, there was no significant change in inequality according to all standard inequality indices. For some specific indices, such as the Gini or the GE(0), there was a statistically significant increase in inequality between 1995 and 2004; for more top-sensitive indices such as GE(1) and GE(2), changes did not differ significantly from zero.

Can these results be reconciled with the evidence of an increase in inequality since the mid-1990s found by researchers using Internal Revenue Service administrative record data on personal adjusted income, notably Piketty and Saez (2003)? In one sense, they cannot, because we have used inequality measures that use information about all incomes in the distribution, ranging from poorest to richest. This is an important advantage of data from large general population surveys such as the CPS. By contrast, the nature of Piketty and Saez's (2003) data means that they focus exclusively on top income shares as their inequality measures. In another sense, however, our results can be reconciled with Piketty and Saez's. Burkhauser, Feng, Jenkins and Larrimore (2008) use exactly the same multiply imputed partially synthetic data as discussed in this paper and consider statistics summarizing top income shares. The estimates of trends in the income shares of subgroups within the richest tenth of the distribution match the Piketty and Saez (2003) estimates, with the exception of the trends for the top 1% of the distribution. Arguably the mismatch in estimated trends for the very richest incomes reflects differences between the data sources in income definitions, sample coverage, and changes over time in the way income is reported. For further details, see Burkhauser, Feng, Jenkins and Larrimore (2008).

Our analysis enables assessment of the public use CPS data augmented by the Census Bureau's cell mean imputations for topcoded observations, at least in the context of estimation of income inequality and its trends. Taking the estimates from the multiply imputed internal data as the gold standard shows that the cell mean-augumented data track

trends in inequality indices over the decade since 1995 reasonably well, though inequality levels are – unsurprisingly – underestimated. However, we have also shown that suppression of income dispersion within cells, combined with use of right censored CPS internal data to construct cell means, also has impacts on variance estimates. Compared to their multiply imputed internal data counterparts, they are underestimated, leading to confidence intervals that are too narrow and a tendency to incorrectly find statistically significant inequality differences. Put another way, cell mean imputations for topcoded observations may do an excellent job of helping estimate mean incomes, but their very nature makes them less suitable for estimation and inference concerning statistics based on higher order moments.

Although few researchers would find it practical to go through the procedures required to access the internal CPS data, and to undertake the research using the data within a U.S. Census Bureau Data Center, we have shown that there is a feasible alternative that works reasonably well. That is, our comparisons of the three series shows a multiple imputation approach applied to topcoded public use CPS data can yield results about income inequality that in several senses lie between those derived using multiple imputation applied to internal data and those derived using cell-mean augmented public use data. The Public-MI approach takes account of income dispersion at the top of the distribution and also takes account of the variability of estimates. Because the public use data are, by definition, in the public domain, it would also be easier for researchers to build more sophisticated imputation models and improve the quality of estimates derived. These models might allow for subgroup differences, for instance allowing for covariates in the estimation of a parametric model or also incorporating the information available from the cell mean imputations. As argued in the Introduction, the multiple imputation methods we propose may be applied in a number of contexts beyond the CPS: right censoring is a relatively common feature of income data sets.

REFERENCES

An, D., and Little, R. J. A. (2007), "Multiple Imputation: an Alternative to Top Coding for Statistical Disclosure Control," *Journal of the Royal Statistical Society*, Ser. A, 170, 923–940.

Angle, J. (2003), "The Salamander: A Model of the Right Tail of the Wage Distribution Truncated by Topcoding," Paper presented at the 2003 Conference of the Federal Committee on Statistical Methodology. http://www.fcsm.gov/03papers/Angle_Final.pdf.

Angle, J., and Tolbert, C. (1999), "Topcodes and the Great U-Turn in Nonmetro/Metro Wage and Salary Income Inequality," Staff Report #9904, Washington, DC: Economic Research Service, U.S. Department of Agriculture.

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

Atkinson, A. B., Rainwater, L., and Smeeding, T. M. (1995), *Income Distribution in OECD Countries. Evidence from the Luxembourg Income Study*, Social Policy Studies No. 18, Paris: Organization for Economic Cooperation and Development.

Atkinson, A. B., and Piketty, T. (eds.) (2007), *Top Incomes over the Twentieth Century A Contrast between European and English-Speaking Countries*, Oxford: Oxford University Press.

Autor, D., Katz, L., and Kearney, M. (2008), "Trends in U.S. Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics*, 90, 300–323.

Bandourian, R., McDonald, J. B., and Turley, R. S. (2003), "A Comparison of Parametric Models of Income Distribution across Countries and Over Time," *Estadistica*, 55, 135– 152.

Beach, C. M., and Richmond, J., (1985), "Joint Confidence Intervals for Income Shares and Lorenz Curves," *International Economic Review*, 26, 439–450.

Bernstein, J., and Mishel, L. (1997), "Has Wage Inequality Stopped Growing?" *Monthly Labor Review*, 120, 3–16.

Biewen, M., and Jenkins, S. P. (2005), "svygei_svyatk: Stata Modules to Derive the Sampling Variances of Generalized Entropy and Atkinson Inequality Indices When Estimated From Complex Survey Data," Statistical Software Components Archive S453601. http://ideas.repec.org/c/boc/bocode/s453601.html

Biewen, M., and Jenkins, S. P. (2006), "Variance Estimation for Generalized Entropy and Atkinson Inequality Indices: the Complex Survey Data Case," *Oxford Bulletin of Economics and Statistics*, 68, 371–383.

Bishop, J. A., Chiou, J.-R., and Formby, J. P. (1994), "Truncation Bias and the Ordinal Evaluation of Income Inequality," *Journal of Business and Economic Statistics*, 12, 123–127.

Bishop, J. A., Formby, J. P., and Smith, W. J. (1991*a*), "International Comparisons of Income Inequality: Tests for Lorenz Dominance Across Nine Countries," *Economica*, 58, 461–477.

Bishop, J. A., Formby, J. P., and Smith, W. J. (1991*b*), "Lorenz Dominance and Welfare: Changes in the U.S. Distribution of Income, 1967–1986," *Review of Economics and Statistics*, 73, 134–139.

Bordley, R. F., McDonald, J. B., and Mantrala, A. (1996), "Something New, Something Old: Parametric Models for the Size Distribution of Income," *Journal of Income Distribution*, 6, 91–103.

Brachmann, K., Stich, A., and Trede, M. (1996), "Evaluating Parametric Income Distribution Models," *Allgemeines Statistisches Archiv*, 80, 285–298.

Burkhauser, R. V., Feng, S., and Jenkins, S. P. (in press), "Using the P90/P10 Ratio to Measure US Inequality Trends with Current Population Survey Data: A View from Inside the Census Bureau Vaults," *Review of Income and Wealth*.

Burkhauser, R. V., Feng, S., Jenkins, S. P. and Larrimore, J. (2008), "Estimating Trends in US Income Inequality Using the Current Population Survey: The Importance of Controlling for Censoring," ISER Working Paper 2007-25, Colchester, U.K.: Institute for Social and Economic Research, University of Essex. http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2008-25.pdf

Burkhauser, R. V., Feng, S., and Larrimore, J. (2008), "Measuring Labor Earnings Inequality Using Public-Use March Current Population Survey Data: The Value of Including Variances and Cell Means When Imputing Topcoded Values," NBER Working Paper 14458, Cambridge MA: National Bureau of Economic Research. http://www.nber.org/papers/w14458

Burkhauser, R. V., and Larrimore, J. (in press), "Trends in the Relative Economic Well Being of Working-Age Men with Disabilities: Correcting the Record using Internal Current Population Survey Data," *Journal of Disability Policy Studies.*

Cowell, F. A. (2000), "Measurement of Inequality," in *Handbook of Income Distribution*, (eds) A. B. Atkinson and F. Bourguignon, Amsterdam: Elsevier North Holland, pp. 87–166.

Cowell, F. A., and Victoria-Feser, M.-P. (1996), "Robustness Properties of Inequality Measures," *Econometrica*, 64, 77–101.

Danziger, S., and Gottschalk, P. (1995), *America Unequal*, Cambridge MA: Harvard University Press for the Russell Sage Foundation.

Feng, S., Burkhauser, R. V., and Butler, J. S. (2006), "Levels and Long-Term Trends in Earnings Inequality: Overcoming Current Population Survey Censoring Problems Using the GB2 Distribution," *Journal of Business and Economic Statistics*, 24, 57–62.

Fichtenbaum, R., and Shahidi, H. (1988), "Truncation Bias and the Measurement of Income Inequality," *Journal of Business and Economic Statistics*, 6, 335–337.

Foster, J. E. (1985), "Inequality Measurement," in *Fair Allocation, Proceedings of Symposia in Applied Mathematics* (vol. 33), ed. H. P. Young, Providence, RI: American Mathematical Society, pp. 31–68.

Gartner, H., and Rässler, S. (2005), "Analyzing the Changing Gender Wage Gap Based on Multiply Imputed Right Censored Wages," IAB Discussion Paper No. 5/2005, Nuremberg, Germany: Institute for Employment Research. http://doku.iab.de/discussionpapers/2005/dp0505.pdf

Gottschalk, P., and Danziger, S. (2005), "Inequality of Wage Rates, Earnings and Family Income in the United States, 1975–2002," *Review of Income and Wealth*, 51, 231–254.

Jenkins, S. P. (2006), "svylorenz: Stata Module to Derive Distribution-Free Variance Estimates From Complex Survey Data of Quantile Group Shares of a Total, and Cumulative Quantile Group Shares," Statistical Software Components Archive S456602. http://ideas.repec.org/c/boc/bocode/s456602.html

Jenkins, S. P. (in press), "Distributionally-Sensitive Inequality Indices and the GB2 Income Distribution," *Review of Income and Wealth*.

Katz, L. F. and Murphy, K. M. (1992), "Changes in Relative Wages, 1963-87: Supply and Demand Factors," *Quarterly Journal of Economics*, 107, 35–78.

Kleiber, C., and Kotz, S. (2003), *Statistical Size Distributions in Economics and Actuarial Sciences*, Hoboken, NJ: John Wiley.

Kovačević, M. S., and Binder, D. A. (1997). "Variance Estimation for Measures of Income Inequality and Polarization," *Journal of Official Statistics*, 13, 41–58.

Larrimore, J., Burkhauser, R. V., Feng, S. and Zayatz, L. (2008), "Consistent Cell Means for Topcoded Incomes in the Public Use March CPS (1976–2007)," *Journal of Economic and Social Measurement*, 33, 89–122.

Lemieux, T. (2006), "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *American Economic Review*, 96, 461–498.

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

Piketty, T., and Saez, E. (2003), "Income Inequality in the United States, 1913–1998," *Quarterly Journal of Economics*, 118, 1–39.

Reiter, J. P. (2003), "Inference for Partially Synthetic, Public Use Microdata Sets," *Survey Methodology*, 29, 181–188.

Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley.

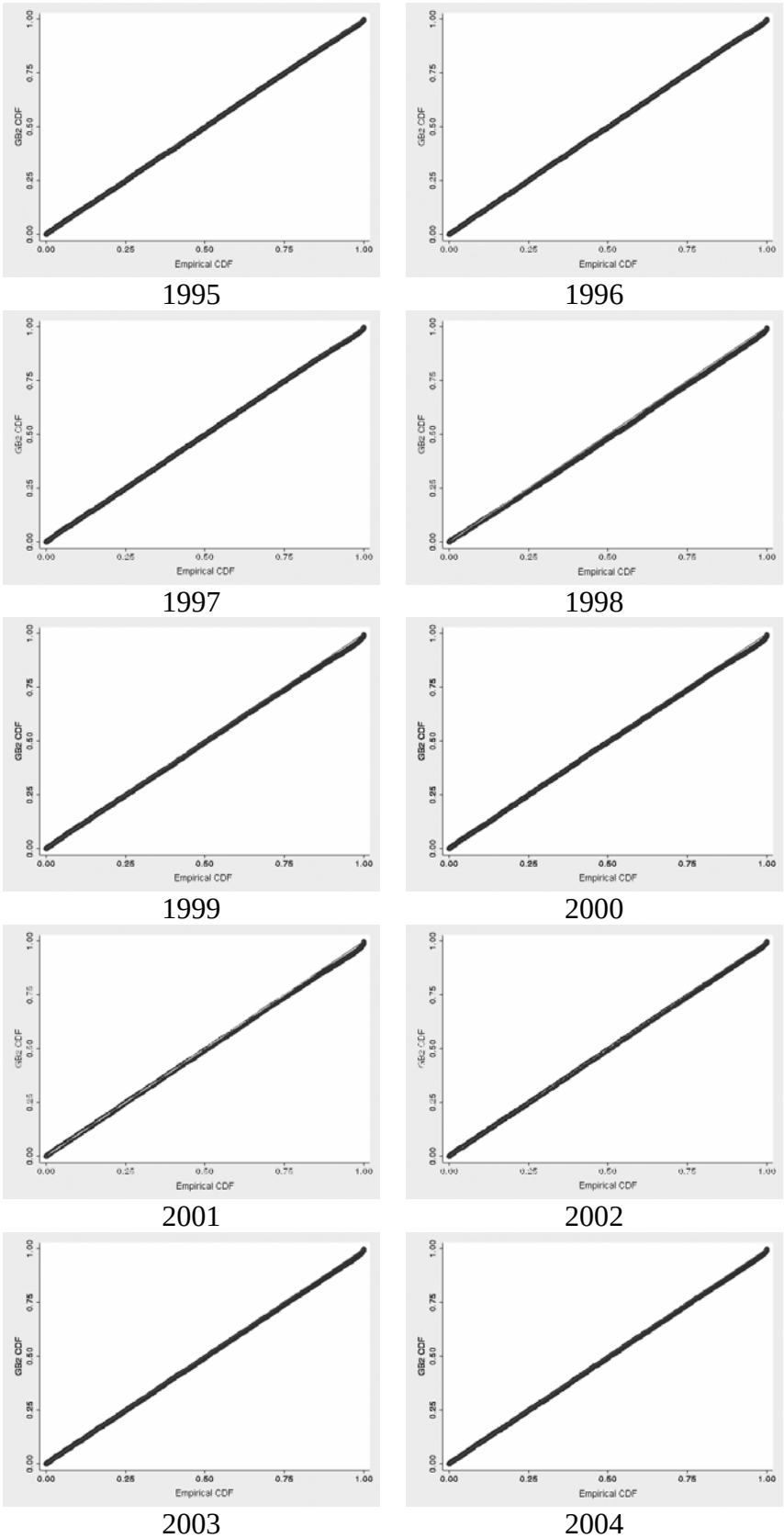
Ryscavage, P. (1995), "A Surge in Growing Income Inequality?" *Monthly Labor Review*, August, 51–61.

Schluter, C., and Trede, M. (2002), "Tails of Lorenz Curves," *Journal of Econometrics*, 109, 151–166.

StataCorp. (2005), *Stata Statistical Software: Release 9*, College Station, TX: StataCorp LP.

U.S. Census Bureau (Various years), *Income, Poverty, and Health Insurance Coverage in the United States: various years*. (August), Current Population Reports, Consumer Income, Washington DC: GPO.

APPENDIX. Probability Plots from GB2 Distribution fit to Public-Use CPS Data



1995 1996

1997 1998

1999 2000

2001 2002

2003 2004

Figure A. Probability plots for GB2 estimates: fitted versus observed. Plot based on richest 70% of each distribution only. GB2 estimates account for left-truncation and right-censoring (see text). Source: authors' derivations from public-use CPS data.