

ISSC DISCUSSION PAPER SERIES

SELECTION BIAS AND MEASURES OF INEQUALITY

*Dr. Rosalia Vasquez Alvarez
Prof. Bertrand Melenberg
Prof. Arthur van Soest*

Rosalia Vasquez is a member of the Economics Dept. in UCD and an associate of the Institute for the Study of Social Change (ISSC). Bertrand Melenberg is a research fellow at CentER, Tilburg University, the Netherlands. Arthur van Soest is a Professor in Econometrics in Tilburg University, the Netherlands.

This paper is produced as part of the Policy Evaluation Programme at ISSC; however the views expressed here do not necessarily reflect those of ISSC. All errors and omissions remain those of the author.

Selection bias and measures of inequality

Rosalia Vazquez Alvarez*

(University College Dublin and ISSC)

Bertrand Melenberg & Arthur van Soest†

(Tilburg University)

* University College Dublin, ISSC

† Tilburg University, Department of Econometrics

(This version: February 2002)

Abstract

Variables typically used to measure inequality (e.g., wage earnings, household income or expenditure), are often plagued by nonrandom item nonresponse. Ignoring non-respondents or making (often untestable) assumptions on the nonresponse sub-population can lead to selection bias on estimates of inequality. This paper draws on the approach by Manski (1989,1994) to derive bounding intervals on both the Gini coefficient and the Inter-Quartile range. Both sets of bounds provide alternative measures of inequality which allow for any type of selective nonresponse, while making no assumptions on the behaviour of non-respondents. The theory is illustrated measuring earnings inequality (over time and between samples) for post-unification Germany over the nineties.

Key words: Selection bias, sample nonresponse, measures of inequality, nonparametric bounds and identification.

JEL Classification: C13, D31, D63, C14

We are grateful to Rob Alessie for helpful comments. Gauss programs used in this papers are available from the corresponding author. Please address correspondence to *Rosalia Vazquez-Alvarez, ISSC, University College Dublin, Belfield Campus, D-4, Ireland. † Department of Econometrics, Tilburg University, Warandelaan 2, Tilburg LE 5000, The Netherlands.

1 Introduction

Micro-economic variables from household surveys are often subject to the problem of missing data. Item nonresponse, as a particular type of missing data, is usually associated with questions that aim at eliciting information in the form of exact amounts from respondents in the sample, so that while individuals surveyed are willing and able to disclose details on family composition, labour market status, etc., a non-negligible percentage of the sample will provide no information on variables such as earnings, total income, savings, or consumption expenditure; these variables are often used to estimate inequality measures, such as Gini coefficient, Theil's coefficient or Atkinson's measure of inequality. Juster et al.(1997) motivates the possibility that cognitive factors (e.g., lack of accurate information or confidentiality reasons on behalf of the respondent) are key elements in explaining why many people are reluctant to disclose information on these type of variables. This implies that non-respondents might not be a random sample, and leads to potential selection problems, since the remaining full respondents may not be a representative sample from the population under study, which translates into the possibility that, in the presence of item nonresponse, inference drawn from estimates of inequality based only on full respondents may lead to biased conclusions if applied to the underlying population.

Traditional approaches to deal with the selection problem range from the assumption of exogeneity (i.e., random nonresponse) to the use of selectivity models (i.e., a joint model of response behavior and the variable of interest, conditional on covariates), or imputation procedures for the missing values, such as hot-deck or multiple imputation.

Both selectivity models and imputation use all information available in the sample to obtain a full set of data, while avoiding the assumption of random nonresponse; this complete set can be used to estimate traditional measures of inequality. However, both procedures share the problem of requiring additional (often untestable) assumptions. For example, selectivity models require alternative assumptions such as those of single index or independence between covariates and error term, while imputation needs to use respondents as a pool of donors for missing information on non-respondents, thus the need to assume that these two populations do not differ in their behavior, conditional on a set of shared characteristics.

Since the early 1990's Charles Manski has put forward a new approach to deal with censored data in the form of item nonresponse that avoids such assumptions; see Manski (1989, 1990, 1994, 1995) and also Heckman (1990). The idea is to use nonparametrics, imposing no assumptions - or much weaker assumptions than those in the parametric or semi-parametric literature -, together with the concept of identification up to a bounding interval. Allowing for any type of nonrandom response behavior, Manski (1989) shows how to derive an upper and lower bound around the parameter of interest, which is usually the distribution function, its

quantiles or functions of its quantiles. The precision with which the parameter of interest is determined, i.e., the width between the upper and the lower bound, depends on the nonresponse probability.

The purpose of this paper is twofold. First, bounds on the Gini coefficient¹ are derived which allow for any type of response behaviour; these bounds prove to be too wide given that it is impossible to tighten the Gini coefficient's natural upper bound. Second, the paper applies the approach by Manski to derive sharp Manski-type bounds around the inter-quartile range (IQR), as an alternative method to measure inequality in the presence of sample selection due to item nonresponse. Analyzing movements in estimates of these latter bounds, joint with estimates of shifts in bounding interval on the quantiles of the distribution, provide an adequate set of tools powerful enough to assess inequality changes between samples and across time.

The theory is illustrated using the variable earnings from the 1990 and 1997 waves of the German Socio Economic Panel (GSOEP). The illustration shows that bounding the Gini coefficient, thus allowing for random (earnings) nonresponse, leads to estimated bounds which are too wide to be useful for empirical work. This illustration also shows that estimating the Gini under alternative assumptions on the nonresponse sub-population leads to results which are not conclusive with respect to inequality trends, unless one makes the (rather strong) assumption that the distribution among respondents and non-respondents remains constant over time. On the other hand, estimates of bounds on the IQR turn out to be much narrower, and thus more informative on the changing inequality trends. The illustration shows that estimates of sharp worst case bounds on IQR, in combination with bounds on the quantiles, appear to be attractive tools to assess changes in earnings differentials and earnings inequality, both for comparisons between populations and for analyzing the trend in inequality in a given population over time.

The remainder of the paper is organized as follows. Section 2 elaborates on the problems associated with measuring inequality in the presence of selection bias due to missing data. Section 3 presents the theoretical framework, first deriving bounds for the Gini coefficient which allow for any type of response behavior, and second deriving a set of sharp worst case bounds for the IQR following the approach by Manski (1989, 1990). Section 4 describes the GSOEP data used in the empirical section. Section 5 presents the empirical results. Section 6 concludes.

¹Slottje et al. (1989) show that many of the traditional inequality measures are theoretically equivalent, including the relation between the Gini coefficient, Atkinson's measure of inequality and a variant of Theil's coefficient. Therefore, although the Gini coefficient is used throughout this paper as the benchmark inequality measures, most of what is said will also applies to other conventional measures of inequality.

2 Selection bias and Inequality measures

The study of income inequality at the micro-economic level requires income data - either at the household or at the individual level - representative of the population under study, which can then be used to estimate well established inequality measures such as the Gini Coefficient, the Theil Coefficient, or Atkinson's measure of inequality. A common feature between these measures is their reliance on some weighted summary of the difference between individual incomes in the data. For example, the Gini Coefficient estimates the average difference between all possible pairs of income in the population, expressed as proportion of total income; if the difference between individual incomes is low, the estimate of the Gini coefficient will be close to zero, indicating low inequality, whereas if total income is concentrated among a few very rich individuals, the estimate will approximate one. Similarly, Theil's inequality measure computes the average difference between the log of each individual income and the mean of income, whereas Atkinson's measure computes the mean of transformed incomes, corresponding to some measure of social welfare (see Cowell (2000) for an extensive discussion of these and other measures of income inequality). In all cases, in order to attain unbiased estimates of these measures, their definition implies the requirement of a complete sample representative of the underlying population. However, it is a well known fact that in household surveys, variables such as income and earnings are typically subject to the problem of item nonresponse. In general, item nonresponse is associated with variables that require the disclosure of an exact amount (other typical examples are consumption, savings, value of assets and debt) since confidentiality problems and/or uncertain knowledge of the amount in question might lead to a "don't know" or "refuse" answer (see Jacowitz et al. (1995) for psychological explanations). The econometric problem with this type of missing values is that the response behavior may be nonrandom, so that those who respond may not constitute a representative sample of the population of interest, and application of standard procedure to the full response sample, while ignoring non-respondents, may result in biased estimates of the parameter of interest, in this case, biased estimates of income inequality. For example, one could estimate the Gini coefficient using the sample of full respondents only, thus allowing for random nonresponse, a practice known as the exogeneity assumption. But if nonresponse is nonrandom, for example, if non-respondents are in the tails of the income distribution, income inequality will be underestimated. The opposite is true if all non-respondents earn the mean income. Likewise, if respondent's behavior changes over time, estimating inequality while assuming that the behaviour of non-respondents is constant between time periods may lead to biased conclusions on the changing trends of income inequality.

The fact that ignoring non-respondents may lead to selection bias has been well

established since the late 1970's, particularly since the seminal work by Heckman (see Heckman (1979), for example). Since then a huge literature has emerged providing parametric and semi-parametric models to deal with selection bias. Selection models postulate the response mechanism jointly with an outcome equation for the dependent variable of interest. The estimates of the outcome equation can be used to impute the missing observations, thus providing a full-response data set that can then be used to estimate income inequality in the standard way. For example, Biewen (1999) applies a selection model to estimate inequality in gross earnings in Germany using the 1997 wave of the GSOEP. He compares the results to those based upon other assumptions, such as exogenous nonresponse. His findings suggest that earnings nonresponse is a u-shaped function of income: nonresponse is higher for both low and high income earners than for intermediate income groups.

Another method to deal with nonresponse is direct imputation. Biewen (1999) applies this method, with imputation based on the matching procedure of Rosenbaum (1995). The method is similar to the hot deck imputation in that both look for individuals in the full response sample who match the characteristics of individuals in the non-response sample. A more complex imputation process is multiple imputation as suggested by Rubin (1997), where each missing income value could be replaced by two or more acceptable values representing a distribution of possibilities.

Clearly, there exist many different techniques to deal with nonresponse; nowadays, assuming that nonresponse is completely exogenous is a rare practice. Still, all these techniques require additional assumptions. Semi-parametric selection models relax these additional assumptions to some extent, but still rely on additional (partly untestable) assumptions, such as exclusion restrictions. Most imputation procedures require that non-response is random conditional on a set of observed covariates. If the additional assumptions are not satisfied, the estimates of inequality measures may be inconsistent, hampering a comparison of inequality in different time periods or between groups. Grabka et al. (1999) and Schwarze (1996) provide examples of this. In both of these studies Theil's $I(0)$ inequality measure is computed using the GSOEP to analyze income inequality between and within East and West Germany over time. The two papers cover different time periods, but use similar techniques. In both studies a composite income variable at the individual level is used. Nonresponse affect each of the income items that make up the final composite variable, so that the final percentage of nonresponse would be too large if all the observations with some nonresponse were deleted. Instead, the missing items of income are imputed using the mean value of the full response sub-sample per income item. But if, for example, non-respondents are typically high and low income earners, this will lead to underestimation of income inequality.

The ideal solution would be to find a method to estimate income inequality that allows

for nonrandom item nonresponse, using all the information available in the data, including the nonresponse sub-sample, and avoiding assumptions on response behavior that cannot be tested. The procedures discussed above do not satisfy these criteria. Since the early 1990's Manski has put forward a new approach to deal with censored data in the form of item nonresponse: see Manski (1989, 1990, 1994 and 1995), but also Heckman (1990). Until now, the main applications of this approach are to be found in the literature on treatment effect - see for example Lechner (1999) and Ginther (1997). The basis for this approach is to use nonparametrics, imposing no assumptions, or much weaker assumptions than the parametric or semi-parametric literature, together with the concept of identification up to a bounding interval. Manski (1989) shows that, without additional assumptions, the sampling process fails to fully identify most features of the conditional distribution of Y given X , but that in many cases a lower bound and an upper bound for the feature of interest (for example, distribution functions of Y given X , its quantiles or functions of its quantiles) can be derived. Manski (1994, 1995) calls these bounds 'worst case bounds' and shows how they can be tightened by adding weak (data) assumptions, such as the assumptions of monotonicity and/or exclusion restrictions (see Vazquez-Alvarez et al. (1999) for an application to the distribution of income in a Dutch cross-section). Thus, Manski's approach lies, not on estimates of the parameter of interest and corresponding confidence bands based on sampling error, but instead proposes to estimate a bounding interval for the parameter of interest which accounts for both sampling error and error due to nonresponse. The resulting bounds solve the selective nonresponse problem at the expense of increasing uncertainty. The advantage is that the identification region contains the population parameter with probability one; for moderate to low levels of nonresponse the bounding interval have shown to be very useful tools for testing economic hypotheses of interest (for example, see Manski (1995) and Vazquez-Alvarez et al. (2000)). In the next section, we review Manski (1989) to derive sharp bounding intervals around the inter-quartile range that allow for item nonresponse to be nonrandom; a joint analysis of such bounds and shifts in the quantiles of the distribution provide an adequate set of tools powerful enough to assess inequality changes between samples and across time. Before deriving bounding intervals on the IQR, the theory section opens by deriving bounds on the Gini coefficient which allow for nonrandom nonresponse; this derivation will show that such bounds may not be very informative, since it is not possible to improve upon the natural upper bound for the Gini coefficient: this will reinforce the use of bounds on the IQR as an alternative measure of inequality trends in the presence of nonrandom item nonresponse.

3 Theoretical framework

3.1 Bounds on the Gini coefficient

This sub-section derives sharp (worst case) bounds for the Gini coefficient which allow for nonresponse to be nonrandom. Let y be the variable of interest, for example, earnings or total income. The sample analogue of the Gini coefficient is given by

$$G_n = \frac{1}{2n^2\bar{y}} \sum_{i=1}^n \sum_{j=1}^n |y_i - y_j| \quad (1)$$

where \bar{y} stands for the complete sample mean and n is the size of the complete sample. The Gini coefficient is bounded in the $[0,1]$ interval. If all values of y are similar, the estimate of (1) will approximate 0, whereas if a large percentage of $\sum_{i \in n} y_i$ is concentrated among a few very rich, the estimate of (1) will approximate 1.

Assume a finite population and let n stand for the size of such (finite) population, with $n = n_1 + n_2$, such that n_1 and n_2 denote the sizes of the full response and the nonresponse populations, respectively. Thus a suffix 1 implies full respondents and 2 non-respondents. Let μ_1 and μ_2 be the population means of the income variable of interest for the two sub-populations and define $p_1 = n_1/n$ and $p_2 = n_2/n$. With this expressions (1) can be re-written as

$$G_n = \frac{p_1^2 G_1 \mu_1 + p_2^2 G_2 \mu_2 + p_1 p_2 \frac{1}{n_1 n_2} \sum_i \sum_j |y_{1i} - Y_{2j}|}{p_1 \mu_1 + p_2 \mu_2} \quad (2)$$

where G_1 and G_2 are the Gini coefficients associated with the full response and nonresponse sub-populations respectively. In the presence of nonrandom nonresponse on y , the Gini coefficient in (2) cannot be identified - since G_2 is not identified either -, unless one makes strong (often untestable) assumptions on the nonresponse sub-population. Doing away with such assumptions, the only information on G_2 is that $0 \leq G_2 \leq 1$. Moreover,

$$\sum_i \sum_j |y_{1i} - y_{2j}| \geq \sum_i \sum_j |n_2 y_{1i} - \sum_j y_{2j}| = n_2 \sum_i |y_{1i} - \mu_2| \quad (3)$$

where the unknown is the mean value of the nonresponse population with $\mu_2 \in [0, \infty)$. The value that minimizes the right hand side of (3) allows for a minimum on (2) which either improves or equals its natural lower bound of 0. Thus, the sharp lower bound on (2) is such that

$$G_n \geq \inf_{m_2 \in [0, \infty)} \frac{p_1^2 G_1 \mu_1 + p_1 p_2 \frac{1}{n_1} \sum_i |y_{i1} - m_2|}{p_1 \mu_1 + p_2 m_2} \quad (4)$$

It is easy to show that the (natural) maximum on the Gini coefficient equals $[(n-1)/n]$. Therefore, for large populations, the sharp upper bound for the Gini coefficient tends to 1, since

$$\frac{n-1}{n} \rightarrow 1 \quad \text{as} \quad n \rightarrow \infty \quad (5)$$

With expressions (4) and (5), bounds on the Gini coefficient are given by,

$$\inf_{m_2 \in [0, \infty)} \frac{p_1^2 G_1 \mu_1 + p_1 p_2 \frac{1}{n_1} \sum_i |y_{i1} - m_2|}{p_1 \mu_1 + p_2 \mu_2} \leq G_n \leq 1 \quad (6)$$

thus allowing for a measure of the Gini coefficient which accounts for any type of nonresponse behavior.

3.2 Bounds around the inter-quartile range (IQR)

Drawing from Manski (1989), this sub-section derives sharp bounds on the inter-quartile range. To do this, we first show how to derive bounds on $P(Y \leq y | \mathbf{x})$, the conditional distribution function of income at a given $y \in \mathbb{R}$, and given $\mathbf{x} \in \mathbb{R}^P$.² Bounds on the (conditional) distribution function can be used to derive bounds on quantiles or functions of quantiles, for example, bounds on the IQR. Let δ be a binary random variable that takes the value of 1 if income is observed, and zero otherwise. With this, $P(Y \leq y | \mathbf{x})$, the population's conditional distribution function given $X = \mathbf{x}$, can be written as:

$$P(Y \leq y | \mathbf{x}) = P(Y \leq y | \mathbf{x}, \delta = 1)P(\delta = 1 | \mathbf{x}) + P(Y \leq y | \mathbf{x}, \delta = 0)P(\delta = 0 | \mathbf{x}) \quad (7)$$

² In the empirical illustration, conditioning is only with respect to West and East Germany, so that the role of the conditioning set X is limited, therefore although the theory is exposed allowing for a conditional set, we do not extend the theoretical discussion to the treatment of item nonresponse in X . For a discussion on this see Manski and Horowitz (1998)

The data identifies $P(Y \leq y|x, \delta = 1)$, $P(\delta = 1|x)$ and $P(\delta = 0|x)$, which can be consistently estimated using, if necessary, some nonparametric regression estimator. On the other hand, the data fails to identify $P(Y \leq y|x, \delta = 0)$, the distribution function for the nonresponse sub-population, and therefore $P(Y \leq y|x)$ cannot be identified either. Assuming exogeneity would solve the problem since it imposes that $P(Y \leq y|x, \delta = 0) = P(Y \leq y|x, \delta = 1)$. If, on the other hand, no assumptions are made with respect to the relation between response behavior and Y , then all we know about the conditional distribution for non-respondents is that $0 \leq P(Y \leq y|x, \delta = 0) \leq 1$. Applying this to expression (7) leads to the following upper and lower bound on $P(Y \leq y|x)$:

$$P(Y \leq y|x, \delta = 1)P(\delta = 1|x) \leq P(Y \leq y|x) \leq P(Y \leq y|x, \delta = 1)P(\delta = 1|x) + P(\delta = 0|x) \quad (8)$$

Expression (8) shows Manski's worst case bounds. The difference between the upper and lower bound is, for any given value of y , equal to $P(\delta = 0|x)$, the conditional percentage of nonresponse. As with expression (6), the worst case bounds in (8) are sharp in the sense that narrower bounds cannot be obtained without making further assumptions. Manski (1995) shows how nonparametric assumptions of monotonicity and exclusion restrictions can lead to sharper bounds in (8).

Expression (8) can be used to derive worst case bounds on the quantiles of the distribution. Let $\alpha \in (0, 1)$; the α -*quantile* of the conditional distribution of Y , given $X=x$, is the smallest number $q(\alpha, x)$ that satisfies $P(Y \leq y|x) \geq \alpha$:

$$q(\alpha, x) \leq \inf \{y: P(Y \leq y|x) \geq \alpha\} \quad (9)$$

Bounds on the quantiles of the distribution can now be obtained by 'inverting' the bounding intervals in (8), so that the lower bound on the distribution function gives an upper bound on the quantile, and likewise, for the lower bound on the distribution with respect to the upper bounds on the quantiles, thus,

$$\inf \{y: P(Y \leq y|x) \geq \beta_1\} \leq \inf \{y: P(Y \leq y|x) \geq \alpha\} \leq \inf \{y: P(Y \leq y|x) \geq \beta_2\} \quad (10)$$

where $\beta_1 = [(\alpha - P(\delta = 0|x))/P(\delta = 1|x)]$ and $\beta_2 = [\alpha/P(\delta = 1|x)]$. Expression (10) shows worst case bounds on the quantiles of the distribution, which can further be used to derive bounds on the conditional inter-quartile range, $IQR(x)$. Using the notation in (9), the conditional inter-quartile

range is given by³:

$$IQR(x) = q(0.75,x) - q(0.25,x) \quad (11)$$

A straight forward way to obtain bounds on $IQR(x)$ would be to consider each of the quantiles in the right hand side of (11) separately: if we denote the lower and upper bound in (10) by $L(\alpha,x)$ and $U(\alpha,x)$, respectively, this gives the following bounds on $IQR(x)$:

$$L(0.75,x) - U(0.25,x) \leq IQR(x) \leq U(0.75,x) - L(0.25,x) \quad (12)$$

However, Appendix A shows that only the lower bound in (12) is sharp. Manski (1994, footnote 2) already pointed out a similar problem for the difference between two values of the distribution function. To illustrate this, let there be two potential values of Y , $t_0 < t_1$. Manski's point refers to $P(t_0 \leq y \leq t_1 | x) = P(y \leq t_1 | x) - P(y \leq t_0 | x)$, that is, the probability that Y is between two values. Using bounds on $P(y \leq t_1 | x)$ and $P(y \leq t_0 | x)$ as given in (8), and manipulating these, gives the following bounds around $P(t_0 \leq y \leq t_1 | x)$:

$$\begin{aligned} P(t_0 \leq y \leq t_1 | x, \delta = 1)P(\delta = 1 | x) - P(\delta = 0 | x) \\ \leq P(t_0 \leq y \leq t_1 | x) \leq \\ P(t_0 \leq y \leq t_1 | x, \delta = 1)P(\delta = 1 | x) + P(\delta = 0 | x) \end{aligned} \quad (13)$$

It is easy to see that bounds in (13) are not sharp: sharper bounds can be derived by directly considering the probability of interest, $P(y \in (t_0, t_1) | x)$; these are given by

$$P(t_0 \leq y \leq t_1 | x, \delta = 1)P(\delta = 1 | x) \leq P(t_0 \leq y \leq t_1 | x) \leq P(t_0 \leq y \leq t_1 | x, \delta = 1)P(\delta = 1 | x) + P(\delta = 0 | x) \quad (14)$$

It is easy to show that bounding interval in (14) is sharp: the width between the upper and lower

³ The IQR defined here suggests an absolute measure of inequality, with the draw back that it depends on the unit of measurement, increasing if all income increase by the same factor. This can be avoided by considering the IQR of log income rather than income level. This would not change anything in the theoretical analysis, thus, although the theory is presented in IQR defined as absolute measures, the empirical section presents and compares estimates both in absolute and relative terms.

bounds equals the probability of nonresponse, whereas the width between the bounds in (13) was twice as large.

Deriving sharp bounds on the $IQR(x)$ is not as straight forwards as the deriving the bounds in (12). The derivation is given in Appendix A; the result of this derivation shows that, for nonresponse probabilities less than 0.25,⁴ the sharp bounds on $IQR(x)$ are given by:

$$L(0.75,x) - U(0.25,x) \leq IQR(x) \leq \max_{p \in [0.25 - P(\delta=0|x)]} \{U(0.5+p,x) - U(p,x)\} \quad (15)$$

The lower bound in (15) is the same as the lower bound in (12). The upper bound in (15) however, is generally smaller than the one in (12).

Bounds in (6) and bounds in (15) provide competing measures of inequality; both measures are free from any distributional or data assumption on the behavior of nonrespondents, thus allowing for any type of nonrandom item nonresponse.

4 Data

In order to assess the usefulness of the bounds, the theory in Section 3 is illustrated with an empirical example concerning net earnings in Germany after unification. The data comes from the 1990 and 1997 waves of the German Socio Economic Panel (GSOEP). This panel is a micro-economic panel with the first wave starting in 1984, and henceforth every year. In 1990 the panel was extended to cover the new adhered East German States. The aim of the panel is to provide data for the analysis of social, economic and living conditions in Germany, with data representative of the German population at individual and household level. The core questions cover demographics, education, labor market status history, earnings, housing, health, household production and finally, an extensive section on subjective data (for example, satisfaction with life, health expectations, etc.). Apart from the a sample which is designed to be representative of the full German population, the panel also contains specific sub-samples which are representatives of minority groups such as foreigners (those who are German residents but of Spanish, Turkish, Italian and Yugoslav origin), and a representative sub-sample of those immigrants who have settled in Germany in recent years.

Interviews are carried out face to face. All members of the household age 16 and over participate as individual members in the panel, while questions at the household level are

⁴ For larger probabilities of nonresponse, the expression becomes easier. Details of theis are not provided since it does not seem to be a practically relevant case.

answered by the assigned household representative. The initial wave in 1984 consisted of 11,654 respondents from a total of 5,624 households. In 1997, after attrition and refreshment over the 14 years, the total number of interviewed individuals was 12,560 from a total of 6,442 households.

We choose to study the earnings inequality using the 1990 and 1997 waves because they cover a period in Germany immediately after unification, thus testing the usefulness of the bounds while providing an example of interest in economics. The data allows us to study inequality trends over time, but also difference in earning's inequality between independent samples (West versus East Germany). Table 1 shows summary statistics for the two waves.

Table 1: Sample size and Summary Statistics, years 1990 and 1997 (estimates based on net monthly earnings in 1997 Deutsche Marks).

| | 1990 | | | 1997 | | |
|---------------------|--------------------|-----------------|-----------------|--------------------|-----------------|-----------------|
| | Unified Germany | West Germany | East Germany | Unified Germany | West Germany | East Germany |
| Observations | 13,245 | 9,016 | 4,229 | 12,560 | 8,916 | 3,644 |
| Employable | 9,230 | 5,848 | 3,382 | 8,318 | 5,723 | 2,565 |
| Wage Earners | 8,738 | 5,476 | 3,262 | 7,233 | 5,124 | 2,109 |
| Net earnings | | | | | | |
| NRP | 557 (0.0790) | 317 (0.081) | 240 (0.074) | 467 (0.089) | 323 (0.095) | 144 (0.066) |
| Mean (s.d) | 2,140 (1,520) | 2,460 (1,570) | 1,140 (460) | 2,610 (1,590) | 2,730 (1,680) | 2,122 (1,030) |
| Median | 1,920 | 2,300 | 1,120 | 2,400 | 2,500 | 2,000 |
| Min/Max | 0 - 22,700 | 0 - 22,500 | 45 - 4,860 | 40 - 30,000 | 40 - 30,000 | 100 - 10,000 |

Note 1: 'Observations' refers to number of individuals in the sample.

Note 2: 'Employable' refers to number of individuals who are employed or actively searching for work.

Note 3: 'Wage Earners' are those who declare to be employed and earning wages or profits at the time of the survey.

Note 4: 'NRP' stands for 'Number of Nonresponse'. All estimates in the 'Net earnings' row refer to weighted estimates, thus the percentage of Nonresponse shown in brackets reflects population estimates rather than sample estimates.

Note 5: Min/Max shows the minimum and the maximum amounts of net income per sample and per year.

The number of observations refers to the number of potential respondents age 16 or over, per year and sample. We are interested on individuals who declare to be actively participating in wage/salary earning activities; this particular way of defining wage earners also include self-employed and women in maternity leave, but excludes the unemployed. Sample units are selected using a question which is repeated in all waves, where individuals are asked to classify their own employment status. Those who report to be current wage earners, are asked to declare their monthly wages and/or salary, both in gross and net terms. The empirical study is based on the variable net earnings.

To make the analysis representative of the population, the variable net earnings is weighted with the cross-sectional sampling weights provide by the GSOEP data set.⁵ This explains why reported nonresponse rates (NPR) in Table 1 differ from sample nonresponse rates. For example, un-weighted nonresponse rate (as percentage of wage earners) for West Germany in 1990 was 0.064, but once the sample is weighted the population nonresponse rate for the same sample and period increases to 0.079. The values of net earnings are also corrected for inflation, using different consumer price index for each Unified Germany, West Germany and East German: this allows to compare purchasing power inequality between samples and over time. The base is chosen as 1997, since prices fluctuated greatly over the 1990, and only after 1997 there seems to be some form of stabilization for CPI within regions and for Unified Germany (see Appendix B).

The summary statistics for net earning are based upon the full response sample only; they suggest that at the beginning of the 1990's there were huge earnings differentials between East and West. Although this has diminished considerably over time, the difference in 1997 is still substantial with East German salaries significantly below those of West Germans.

5 Results and estimation method

This section illustrates the theory of Section 3 using the variable net earnings as described in Section 4. The results in Section 5.1 shows that if we want to avoid (often strong) untestable assumptions on the nonresponse sub-population, bounding the inter-quartile range becomes a more informative measure of inequality changes than either bounds or point estimates on traditional measure of inequality such as the Gini coefficient. The results in Section 5.2 compares estimates of quantiles of the distribution assuming random nonresponse to estimates where nonresponse is assumed to be nonrandom, and show how in both cases these estimates can be used to test for earnings equality over time and between samples. These conclusions are combined with those obtained from point estimates on the inter-quartile range - if assuming random nonresponse - as well as estimates of sharp bounds on the inter-quartile range - if assuming nonrandom nonresponse -, to test for changing trends on earning's inequality in Germany after unification.

The bounds presented in Section 3 are defined in terms of sub-populations characteristics, and can be estimated using the corresponding sample analogues. In the case of

⁵ Using weights allow us to include the sub-samples of foreigners and new immigrants since unification. The weights also correct for the larger sampling rate in East Germany compared to West Germany.

a set of continuously distributed conditioning variables X , nonparametric regression techniques can be used to estimate expressions such as (8), (12) and (15) (see for example Härdle and Linton (1994)). However, estimates in this illustration are not based on conditioning on continuously distributed variables, thus, there is no need for smoothing; the estimates of the bounds are functions of the sample analogues of populations and sub-population fractions. All this sample analogues are weighted using the sampling weights provided with the GSOEP data, to correct for stratified and non-representative sampling.

All estimates will be presented together with estimated upper and lower confidence bands, resulting from a bootstrap procedure: re-sampling with replacement 500 times from the original sample. This technique yields two sided 95% confidence bands around the estimated bounds, given by the 2.5th and 97.5th percentile in these 500 estimates. Only the upper confidence band for the upper bound and the lower confidence band for the lower bound will be presented. The resulting (pointwise) difference between these two bands shows the imprecision due to finite sampling error as well as error due to nonresponse, while the difference between the point estimates of upper and lower bounds is an estimate of the imprecision due to nonresponse only.

5.1 Estimates of inequality and bounds on inequality measures

Table 2 shows the result of estimating the Gini coefficient - for different samples and across time -, imposing various alternative assumptions on the earnings distribution among non-respondents. Smaller middle numbers - per sample and time period - show point estimates of the Gini coefficient, whereas each pair of blacken numbers are corresponding 95% confidence intervals estimated using the bootstrap method as described in Section 5.

G(exogenous) refers to estimates of the Gini coefficient assuming nonrandom nonresponse, and therefore using full respondents only. G(low) and G(high) refer to estimates of the Gini coefficient when non-respondents are randomly assigned a value from the lower and upper deciles of the full response distribution, respectively. G(mean) implies that non-respondents are all assigned a value of earnings drawn from a normal distribution with mean and variance equal to the sample mean and variance of the full respondents. Finally, G(median) shows the consequence of simply assigning the median of the respondents to the non-respondents. Table 2 shows that the value of the Gini coefficient can vary substantially between alternative assumptions on the earnings of the nonresponse sub-population. If it is assumed that the relation between the distributions among respondents and non-respondents remains the same over time for each sample, i.e., if one specific column - for any given sample - is considered, then the conclusion is the same irrespective of which assumption is made: between 1990 and 1997 inequality in earnings increased significantly for Unified Germany, did not change significantly

for West Germany and increased significantly for East Germany.

Table 2: Estimates of Gini Coefficient under various assumptions on earnings in the nonresponse population (estimates based on net monthly earnings in 1997 Deutsche Marks)

| Unified Germany | G(exogenous) | G(low) | G(high) | G(mean) | G(median) |
|-----------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| 1990 | [0.3529; 0.3609 ;0.3696] | [0.3803; 0.3892 ;0.4003] | [0.3984; 0.4109 ;0.4281] | [0.3538; 0.3618 ;0.3711] | [0.3382; 0.3459 ;0.3552] |
| 1997 | [0.3060; 0.3158 ;0.3261] | [0.3340; 0.3460 ;0.3570] | [0.3485; 0.3629 ;0.3775] | [0.3054; 0.3157 ;0.3293] | [0.2891; 0.2991 ;0.3097] |
| West Germany | G(exogenous) | G(low) | G(high) | G(mean) | G(median) |
| 1990 | [0.3137; 0.3241 ;0.3346] | [0.3449; 0.3559 ;0.3685] | [0.3574; 0.3707 ;0.3903] | [0.3146; 0.3250 ;0.3369] | [0.2980; 0.3082 ;0.3184] |
| 1997 | [0.3099; 0.3214 ;0.3318] | [0.3379; 0.3498 ;0.3624] | [0.3477; 0.3646 ;0.3802] | [0.3086; 0.3209 ;0.3352] | [0.2899; 0.3035 ;0.3147] |
| East Germany | G(exogenous) | G(low) | G(high) | G(mean) | G(median) |
| 1990 | [0.2073; 0.2152 ;0.2223] | [0.2455; 0.2557 ;0.2665] | [0.2428; 0.2591 ;0.2688] | [0.2080; 0.2146 ;0.2241] | [0.1978; 0.2051 ;0.2121] |
| 1997 | [0.2396; 0.2520 ;0.2650] | [0.2696; 0.2857 ;0.3043] | [0.2756; 0.2984 ;0.3176] | [0.2399; 0.2512 ;0.2665] | [0.2294; 0.2415 ;0.2546] |

However, if we relax the assumption of no over time change in the relation between respondents and non-respondents, Table 2 is no longer valid to draw conclusions on earnings inequality trends. For example, if non-respondents in 1990 West Germany are all median wage earners, but in 1997 the composition of non-respondents for the same population changes to be high income earners, then the Gini for the population as a whole would increase from 0.3082 to 0.3646 and such increase would be significant (notice that the 95% confidence intervals for West Germany G(high) in 1997 - [0.3477; 0.3802] -, envelops the analogous confidence intervals for G(median) in 1990 - [0.2980; 0.3184]); this would contradict conclusions based on separate analysis of either G(high) or G(median). Thus, without making the additional assumption of no change on the composition of non-respondents over time, the presence of nonresponse means that estimates of the Gini in Table 5 cannot point towards a particular trend on earnings inequality for any of the three samples.

Table 3 shows estimates of bounding intervals for the Gini coefficient - expression (6), Section 3 -, and the Inter-Quartile range (IQR) - expression (15), Section 3 -, proposing these as alternative measures which allow for selective nonresponse. In the same table, the upper part shows estimates for the IQR assuming random nonresponse, with estimates for G(exogenous) from Table 2 added for comparative reasons.

Table 3: Estimates of Inequality Measures (Gini Coefficients and Inter-Quartile Range), over time and between samples (measures based on net monthly earning in 1997 Deutsche Marks).

| | Unified Germany | | West Germany | | East Germany | |
|--------------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| Random Nonresponse | 1990 | 1997 | 1990 | 1997 | 1990 | 1997 |
| Gini Coefficient | [0.3529; 0.3609 ;0.3696] | [0.3060; 0.3158 ;0.3261] | [0.3137; 0.3241 ;0.3346] | [0.3099; 0.3214 ;0.3318] | [0.2073; 0.2152 ;0.2223] | [0.2396; 0.2520 ;0.2650] |
| Inter-Quartile range (Levels) | [1,490; 1,580 ;1,595] | [1,310; 1,430 ;1,515] | [1,360; 1,485 ;1,600] | [1,410; 1,520 ;1,640] | [400; 420 ;450] | [660; 740 ;800] |
| Inter-Quartile range (Ln) | [0.959; 1.003 ;1.020] | [0.660; 0.723 ;0.758] | [0.687; 0.764 ;0.849] | [0.667; 0.742 ;0.811] | [0.440; 0.468 ;0.494] | [0.453; 0.506 ;0.551] |
| Non-Random Nonresponse | 1990 | 1997 | 1990 | 1997 | 1990 | 1997 |
| Bounds on the Gini estimate | [0.3335; 0.3421 ;1] | [0.2867; 0.2967 ;1] | [0.2960; 0.3059 ;1] | [0.2881; 0.3005 ;1] | [0.1973; 0.2044 ;1] | [0.2284; 0.2406 ;1] |
| Inter-Quartile range (Levels) | [1,330; 1,410; 2,020 ;2,110] | [1,090; 1,230; 1,700 ;1,860] | [1,100; 1,240; 1,730 ;1,810] | [1,170; 1,300; 1,900 ;2,000] | [360; 370; 480 ;500] | [620; 660; 870 ;950] |
| Inter-Quartile range (Ln) | [0.861; 0.895; 1.132 ;1.170] | [0.542; 0.623; 0.990 ;1.070] | [0.560; 0.630; 0.995 ;1.055] | [0.560; 0.630; 1.130 ;1.195] | [0.390; 0.410; 0.540 ;0.580] | [0.420; 0.450; 0.590 ;0.670] |

Table 3 shows point estimates and estimated bounding intervals for the IQR in both levels and (natural) logs; this allows to check if a change in earnings inequality is caused by either an absolute or a relative (within sample) changes in real earnings.⁶

⁶ Estimating the IQR in levels and comparing these between periods measures inequality changes in absolute terms; for example, if all individual's earnings increased by a similar percentage over the 1990's, earnings inequality would not change, although an estimate of IQR in levels for 1997 would show an increase relative to that of 1990. On the other hand, IQR

With Table 2 it became evident that trends on earnings inequality using the Gini coefficient were not independent from the assumptions made on the behaviour of the nonresponse population over time. This assumption, as well as any of the initial 5 assumptions in Table 2, can be relaxed if one estimates expression (6) which bounds the Gini coefficient from below. However, Table 3 (lower part) shows that the estimated worst case bounds on the Gini are too wide to draw meaningful information on trends in inequality, since the overlap of these estimates between periods, for any of the three populations considered, implies a wide range of patters in changing inequality trends.

The alternative is to look at estimates of the IQR. Assuming exogeneity (upper part, Table 3), estimates of the IQR(level) for Unified Germany would suggest a very small left shift in the confidence interval for the IQR - thus a possible decrease in earnings inequality - between 1990 and 1997. However, once the assumption of exogeneity is relaxed (lower part, Table 3), the confidence bands corresponding to the bounds on the IQR suggest a substantial overlap between the years 1990 and 1997.⁷ Thus, without making further assumptions on the behaviour of the nonresponse sub-population, the evidence do not suggest a change on earnings inequality over time for Unified Germany. Similar conclusions would apply to estimates for West Germany; either under the assumption of exogeneity or using the worst case bounds, the results for West Germany indicate no change in earnings inequality between 1990 and 1997. For East Germany estimates of IQR(level) and worst case bounds of IQR(level) present evidence of a significant increase on earnings inequality over the period; this is because of a significant rightward shift on the range of values - for either point estimate or estimated worst case bounds - which leaves no overlap in the 95% confidence regions between 1990 and 1997. But such evidence, based on estimates of IQR in levels, are not in accordance with estimates of IQR in (natural) logs; estimates of this latter - either assuming random or non-random item non-response - show 95% confidence bands which overlap between 1990 and 1997, thus, looking at estimates of IQR in (natural) logs for East Germany suggest no change in earnings inequality, as was the case for West Germany or Unified Germany. One explanation for the discrepancy between conclusions

measures in (natural) logs is a relative measure of income inequality and would not change if all incomes changed by a similar factor. Thus, it is important to report the IQR in (natural) logs to check that a change in the IQR at levels between periods is due to a relative (within sample) change in earnings.

⁷ Notice that despite this overlap, the bounds on the IQR are still sufficiently narrow to define a particular pattern on the changing trend on income inequality. This was not the case for estimates of worst case bounds on the Gini coefficient.

based on IQR(level) and IQR(ln) is that real earnings increasing substantially for East Germans over the 1990's, but such increase was at a similar rate for every one, thus the significant change in IQR(level) between 1990 and 1997. At the same time, and despite the increase for all in real earnings, discrepancies between East German's real earnings remained fairly stable over the period, that is, relative earnings inequality did not change. It is possible that IQR(level) is picking up the effect of the massive subsidies that arrived from the West aimed at speeding up the process of unification (by reducing wage differentials between East and West), while IQR(ln) shows that the effect of such subsidies did not have the adverse effect of increasing inequality between East German's wage earners.

5.2 Estimates of Quantiles and Inequality

The findings in Section 5.1 suggest no changes in earnings inequality for either Unified Germany or for East and West Germany as distinct regions. In most cases we would expect that the IQR will lead to similar conclusions about trends on inequality than other inequality measures. But IQR can be insensitive to within quantile changes. A joint study of changes in IQR together with the changing patten in the distribution of earnings over time allows for a better understanding on the changing pattern of earnings inequality. This joint study validate the use of the IQR as a measures of inequality which is flexible enough to allow for selective nonresponse while making no assumptions on the nonresponse sub-population

Table 4 compares the distribution of earnings between 1990 and 1997 using a selection of quantiles for Unified Germany, as well as West and East Germany. Assuming random nonresponse (top rows), the 1990 median in earnings for Unified Germany was estimated between DM 1,810 and DM 1,930. Relaxing this assumption and allowing for any type of nonrandom nonresponse implies using bounds on the unknown quantiles - expression (10), Section 3 -,so that the 1990 median for Unified Germany is now estimated between DM 1,660 and DM 2,120 (bottom rows). The increase in the confidence region illustrates the trade off between using strong distributional assumptions - such as random nonresponse - and increasing uncertainty, since the 95% confidence interval assuming random nonresponse only reflects sampling error, but once this assumption is relaxed, the 95% confidence region on the estimated bounds accounts for both sampling error and error due to nonresponse.

The top rows in Table 4 show that the distribution of earnings for Unified Germany and for East Germany have experience a significant upward shift over time; in both cases we observe that for any given percentile, the 95% upper confidence band for 1990 is always above the corresponding 1997 lower 95% confidence band. This is not true for West Germany; in this case, the same top rows show a significant overlap between the 1990 upper 95% confidence

band and corresponding 1997 lower 95% confidence bands. Thus, estimates assuming random nonresponse during the 1990's suggest an increase in real earnings for Unified Germany driven by an increase in real earnings in East Germany.

Table 4: Estimated 95% Confidence Bands around the Quantiles of the distribution for a selection of quantiles. Compares assumptions of Random and Non-Random Nonresponse (Monthly net earnings in 1997 DM; Samples are West German, East German and Unified Germany).

| | Unified Germany | | West German | | East German | |
|-----------------------------|-----------------|-----------------|----------------|-----------------|-----------------|----------------|
| Random Nonresponse | 1990 | 1997 | 1990 | 1997 | 1990 | 1997 |
| 10 th Percentile | (600 , 660) | (690; 850) | (620; 690) | (660; 820) | (580; 640) | (790; 1,000) |
| 25 th Percentile | (1,020; 1,080) | (1,500; 1,700) | (1,345; 1,570) | (1,490; 1,780) | (850; 880) | (1,490; 1,600) |
| 50 th Percentile | (1,810; 1,930) | (2,300; 2,500) | (2,240; 2,350) | (2,490; 2,600) | (1,100; 1,140) | (2,000; 2,090) |
| 75 th Percentile | (2,720; 2,840) | (3,190; 3,400) | (2,990; 3,130) | (3,400; 3,600) | (1,340; 1,400) | (2,500; 2,670) |
| 90 th Percentile | (3,850; 4,080) | (4,500; 4,780) | (4,150; 4,450) | (4,560; 5,000) | (1,610; 1,700) | (3,100; 3,500) |
| Non-random Nonresponse | 1990 | 1997 | 1990 | 1997 | 1990 | 1997 |
| 10 th Percentile | (220; 680) | (0; 900) | (220; 740) | (0; 890) | (210; 670) | (450; 1,000) |
| 25 th Percentile | (840; 1,130) | (1,030; 1,780) | (1,000; 1,680) | (960; 1,800) | (750; 910) | (1,290; 1,640) |
| 50 th Percentile | (1,660; 2,120) | (2,190; 2,600) | (2,130; 2,460) | (2,300; 2,800) | (1,060 , 1,190) | (1,940; 2,100) |
| 75 th Percentile | (2,610; 3,260) | (3,040; 3,900) | (2,870; 3,580) | (3,190; 4,150) | (1,340; 1,480) | (2,400; 2,900) |
| 90 th Percentile | (3,740; 6,800) | (4,100; 50,000) | (4,030; 7,800) | (4,390; 50,000) | (1,580; 2,390) | (3,090; 4,560) |

However, once the assumption of random nonresponse is relaxed (bottom rows), the estimated region of uncertainty for the unknown quantiles increases at each percentile and for all the samples considered; in this case, only East Germany shows a significant increase on real earnings throughout the income distribution, whereas the null of no change in real earnings over the 1990's for both Unified Germany and West Germany cannot be rejected. It seems as if the policies of subsidizing wage earners in East Germany with funds from the West had the desired effect of reducing wage differentials between regions, while leaving the distribution of earnings almost unchanged for Unified Germany over the 1990's, and at the same time, the flow of funds from West to East meant the real earnings in West Germany did not increase over the period. The upward shift in the earnings in East Germany between 1990 and 1997 is detected for all

quantiles in the distribution. This evidence support the evidence on earnings inequality of Section 5.1: the substantial increase in earning were experienced for all levels of earnings - thus the shifting IQR(level) -; such sift brought the distribution of East Germany wages in line to those of West Germans, but the effect on earnings inequality was not significant - as suggested by the insignificant shift in IQR(ln) in Section 5.1.

As a final exercise, Figures 1 to 4 illustrate how estimated bounding intervals can be used as an informal test for changes in real earnings over time, while allowing for any type of nonrandom nonresponse in earnings.

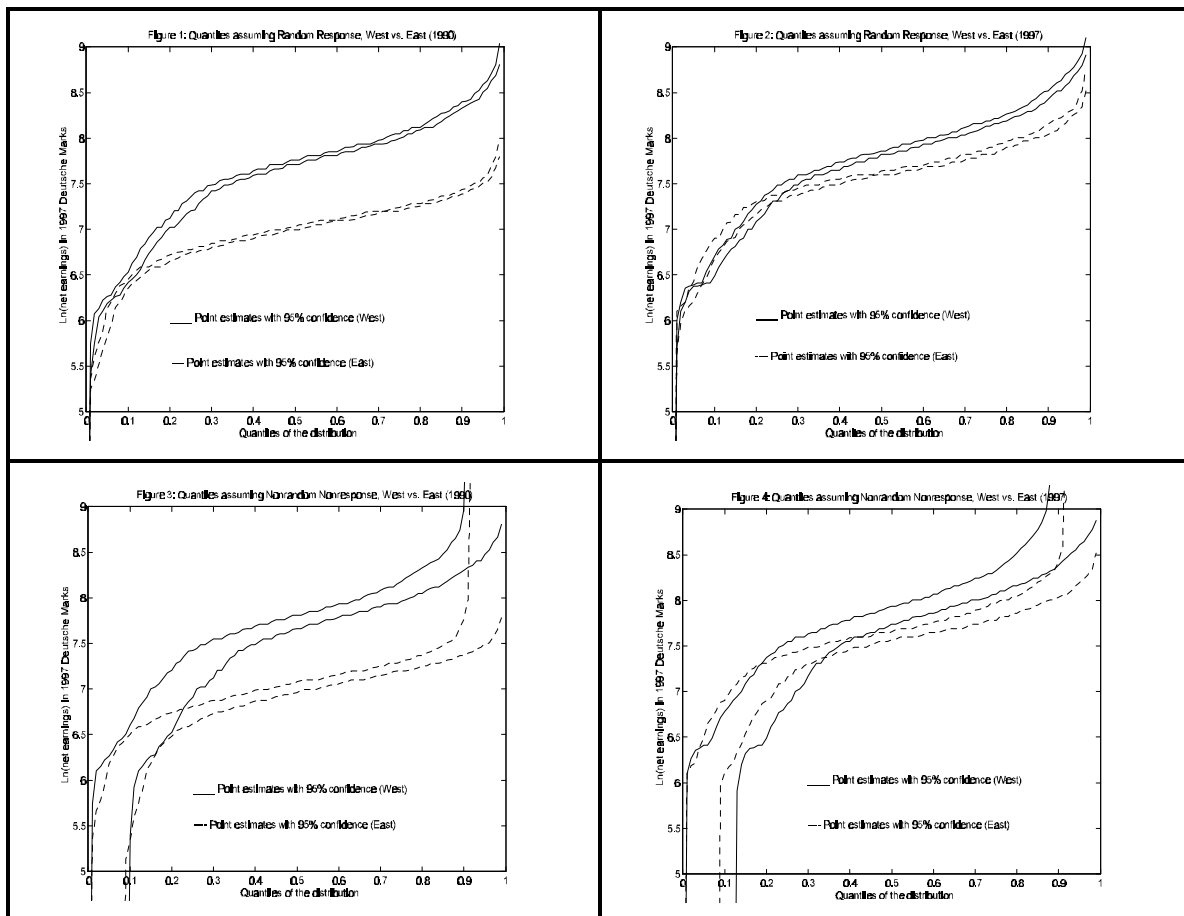


Figure 1 shows estimated 95% confidence bands for the quantiles of the distribution in 1990 for both East and West Germany; Figure 2 shows similar estimates for 1997. In both cases the assumption is that nonresponse is random. Figure 1 shows that in 1990 the null of equality in real earnings between East and West Germany is rejected throughout the earnings distribution. Figure 2 shows the increase in real earnings in East Germany over the 1990's while real earnings for West Germany remained stable over the same period. The resulting overlap between the two regions of confidence suggests that in 1997 the null of equality in real earnings

between East and West can only be rejected for quantiles above the 25th percentile, with some evidence to suggest that in 1997 higher earnings for the lower tail of the distribution in East than in West Germany. In Figures 3 and 4 the assumption of random nonresponse is relaxed, and 95% confidence bands on estimated worst case bounds are presented - expression (10), Section 3 -. In this case, Figure 3 shows that in 1990 the null of earnings equality between East and West cannot be rejected for earners in the tails of the distribution. Comparing Figure 4 to Figure 3 shows that in 1997, once we relax the assumption of random nonresponse, the null of earnings equality between East and West shrinks; only the quantiles between the 40th and 85th percentile of the distribution reject the null of equality in real earnings between East and West Germany. Table 5 provides the formal testing procedure for Figures 2, 3, and 4.⁸ The formal test agrees closely with the informal (graphic) test; allowing for random nonresponse (top rows), in 1997 there was a significant difference between East and West German real earnings for quantiles above the 30th percentile; once we allow for any type of nonrandom nonresponse (bottom rows), the formal test shows that in 1997 real earnings in East Germany were significantly different from those in West Germany for quantiles above the 40th percentile. In both cases, there income differentials between regions has been reduced dramatically and, as suggested by evidence in Section 5.1, such reduction in wage differentials did not result on increase earnings inequality in Unified Germany.

Table 5: Testing for income differentials between independent samples. (Numbers are monthly net

⁸ In each of the three cases, the test statistic is given by the difference of the lower confidence band for West Germany and the upper confidence band for East Germany, divided by the standard error of this difference. The standard error is determined by bootstrapping each of the independent samples, as described in Section 4. The null of equality is rejected if the test statistic exceeds the one tailed critical value of the standard normal distribution.

earnings in 1997 Deutsche Marks)

West vs. East (1997),
Random Non-Response
(cf. Figure 2)

| | Point Estimate, West 1997 (lower bound) | Standard error of point estimate | Point Estimate, East 1997 (upper bound) | Standard error of point estimate | Difference (West 97 - East 97) | Pool standard error | Test Statistic |
|-----------------------------|--|---|--|--|--------------------------------------|---------------------------|-------------------|
| 10 th Percentile | 749 | 40 | 898 | 50 | -149 | 64 | -2.32 |
| 20 th Percentile | 1,296 | 66 | 1,393 | 41 | -97 | 77 | -1.26 |
| 25 th Percentile | 1,660 | 56 | 1,536 | 42 | 124 | 70 | 1.77 |
| 30 th Percentile | 1,899 | 40 | 1,688 | 30 | 211 | 50 | 4.22 |
| 40 th Percentile | 2,193 | 50 | 1,880 | 33 | 313 | 60 | 5.22 |
| 50 th Percentile | 2,499 | 33 | 1,998 | 22 | 501 | 40 | 12.61 |
| 60 th Percentile | 2,848 | 45 | 2,193 | 13 | 655 | 47 | 13.91 |
| 75 th Percentile | 3,498 | 45 | 2,550 | 51 | 948 | 68 | 13.99 |
| 80 th Percentile | 3,720 | 63 | 2,777 | 55 | 943 | 84 | 11.28 |
| 90 th Percentile | 4,797 | 127 | 3,260 | 106 | 1,537 | 165 | 9.32 |

West vs. East (1990), Non-Random
Non-Response
(cf. Figure 3)

| | Point Estimate, West 1997 (lower bound) | Standard error of point estimate | Point Estimate, East 1997 (upper bound) | Standard error of point estimate | Difference (West 97 - East 97) | Pool standard error | Test Statistic |
|-----------------------------|--|---|--|--|--------------------------------------|---------------------------|-------------------|
| 10 th Percentile | 390 | 59 | 626 | 18 | -236 | 61 | -3.84 |
| 20 th Percentile | 785 | 51 | 831 | 14 | -46 | 53 | -0.87 |
| 25 th Percentile | 1,120 | 29 | 881 | 11 | 239 | 31 | 7.64 |
| 30 th Percentile | 1,345 | 59 | 947 | 14 | 398 | 61 | 6.52 |
| 40 th Percentile | 1,905 | 31 | 1,045 | 13 | 860 | 34 | 25.31 |
| 50 th Percentile | 2,180 | 47 | 1,152 | 11 | 1,028 | 48 | 21.33 |
| 60 th Percentile | 2,462 | 14 | 1,270 | 14 | 1,192 | 20 | 60.20 |
| 75 th Percentile | 2,966 | 50 | 1,456 | 7 | 1,510 | 51 | 29.72 |
| 80 th Percentile | 3,250 | 53 | 1,580 | 16 | 1,670 | 56 | 30.07 |
| 90 th Percentile | 4,209 | 70 | 2,065 | 129 | 2,144 | 147 | 14.62 |

West vs. East (1997), Non-Random
Non-Response
(cf. Figure 4)

| | Point Estimate, West 1997 (lower bound) | Standard error of point estimate | Point Estimate, East 1997 (upper bound) | Standard error of point estimate | Difference (West 97 - East 97) | Pool standard error | Test Statistic |
|-----------------------------|--|---|--|--|--------------------------------------|---------------------------|-------------------|
| 10 th Percentile | 280 | 201 | 935 | 47 | -655 | 206 | -3.18 |
| 20 th Percentile | 845 | 75 | 1,450 | 42 | -605 | 86 | -7.01 |
| 25 th Percentile | 1,098 | 84 | 1,599 | 28 | -501 | 88 | -5.67 |
| 30 th Percentile | 1,492 | 94 | 1,700 | 32 | -208 | 99 | -2.09 |
| 40 th Percentile | 1,998 | 41 | 1,899 | 26 | 99 | 49 | 2.02 |
| 50 th Percentile | 2,390 | 48 | 2,098 | 29 | 292 | 56 | 5.24 |
| 60 th Percentile | 2,698 | 56 | 2,299 | 39 | 399 | 68 | 5.88 |
| 75 th Percentile | 3,368 | 76 | 2,795 | 64 | 573 | 100 | 5.75 |
| 80 th Percentile | 3,598 | 66 | 2,999 | 47 | 599 | 81 | 7.42 |
| 90 th Percentile | 4,638 | 132 | 3,998 | 196 | 640 | 236 | 2.72 |

6 Conclusions

Drawing on Manski (1989, 1994, 1995), this paper has derived worst case bounds around the IQR and the Gini coefficient. These two sets of bounds provide competing measures of inequality which allow for any type of nonrandom item nonresponse; both avoid the typical distributional and data assumptions associated with parametric and semi-parametric methods of dealing with selective nonresponse.

The empirical section illustrates the theory using the variable net earnings from the 1990 and 1997 waves of the German Socio Economic Panel (GSOEP). These two waves are chosen because they cover post-unification Germany during the nineties, which allows to test the usefulness of the bounds as measures of inequality (over time and between populations) while providing an example of interest in economics. The estimates of the bounds on both, the Gini coefficient and the IQR, are presented with confidence bands estimated using a bootstrap procedure which samples randomly from the data with replacement. Thus both the imprecision due to nonresponse and finite sample error are taken into account. Population (weighted) nonresponse rates fluctuate around the value of 10% over the period under study, with West Germany typically showing a higher rate than East Germany.

Despite moderate nonresponse rates, estimates of bounds on the Gini coefficient turn out to be too wide to be useful for empirical work; the estimates between periods cover a wide range of patterns and no meaningful conclusions can be drawn for the changing trends on earnings inequality over the period. This suggests that the use of the Gini coefficient would require additional, often untestable assumptions on the nonresponse sub-population. The empirical illustration shows that the Gini coefficient is very sensitive to alternative (mutually exclusive) assumptions on the behaviour of non-respondents, each of which would require the additional (fairly strong) assumption that there is no change in the composition of non-respondents over time. The alternative measure relies on estimated Manski-type of bounds on the IQR. Estimates of these bounds turn up to be much narrower and thus much more informative. The empirical illustration shows how these bounds, in combination with bounds on the quantiles of the distribution, appear to be attractive tools to assess changes on earnings differentials and earnings inequality. With this, the results would suggest evidence of a success for those policies which aimed at reducing the earnings gap between East and West Germany in the early nineties. Subsidies from the West increased earnings at all quantiles of the distribution for East Germany, i.e., comparing estimates of Manski's worst case bounds on earnings between 1990 and 1997 cannot reject the null of a significant upward shift in earnings between these two time periods for East Germany, while similar estimates for West Germany cannot reject the null of earnings equality between 1990 and 1997. Bounds on IQR (natural) log for East Germany suggest that relative (within population) earnings inequality has not changed significantly (the 95% bands overlap substantially between periods); similar results are attained

using bounds on $IQR(\text{level})$ and $IQR(\text{natural})$ log for West Germany, thus suggesting that earnings inequality has remained stable also for West Germany.

In all, estimates of these alternative set of bounds have provided a methodological framework useful to assess changing trends on earning differentials and earning inequalities, in this case, with respect to earnings in post-unification Germany. The bounds use all the information available in the data without the need to make untestable assumptions on the non-response sub-population, thus overcoming the selection problem at the expense of increased uncertainty, that is, the identification region is now composed of both sampling error and error due to nonresponse. The method is elegant, intuitively plausible and extremely flexible, and it works as long as nonresponse rates are moderate. In practice, such bounds can be further tighten by means of weak data assumptions, for example, suggesting a monotonic relation between earnings and nonresponse or using exclusion restrictions in a conditioning set.

References

- Biewen, M. (1999), Item nonresponse and inequality measurement: Evidence from the German Earnings Distribution, Department of Economics, *Ruprecht-Karls-Universität Heidelberg*, Discussion paper no.298.
- Burda, M. (1993), The determinants of East-West German migration; some first results, *European Economic Review* 37 (1993), 452-461, North Holland.
- Cowell, F. (2000), Measuring Inequality, manuscript, 3rd Edition (available via www.lse.ac.uk)
- Grabka, M., Johannes Schwarze, Gert Wagner, (1999), How unification and immigration affected the German income distribution, *European Economic Review*, 43, 867-878.
- Härdle, W., and O. Linton, (1994), Applied nonparametric methods, in: *Handbook of Econometrics IV*, R.F. Engle and D.L. McFadden (eds.), North-Holland, Amsterdam, 2297-2339.
- Heckman, J.J., (1979), Sample selection bias as a specification error, *Econometrica*, 47, 153-161.
- Heckman, J.J., (1990), Varieties of selection bias, *American Economic Review, Papers & Proceedings*, 80, 313-318.
- Jakowitz, K. and D. Kahneman (1995), Measures of anchoring in estimation tasks, *Personality and Social Psychology Bulletin*, 21, 1161-1166.
- Manski, C.F., (1989), Anatomy of the selection problem, *Journal of Human resources*, 24, 343-360.
- Manski, C.F., (1990), Nonparametric bounds on treatment effects, *American Economic Review, Papers & Proceedings*, 80, 319-323.
- Manski, C.F., (1994), The selection problem in: *Advances in Econometrics*, C. Sims (ed.), Cambridge University Press, 143-170.
- Manski, C.F., (1995), Identification problems in the social science, *Harvard University Press*.
- Rubin, D., (1987), Multiple imputation for nonresponse in surveys, *John Wiley & Sons ed*, N.Y.
- Rosenbaum, P.R and D.B. Rubin, (1984), Comment: Estimating the effects caused by treatment, *Journal of the American Statistical Association*, 79, 26-28.
- Schwarze, J. (1996), How income inequality changed in Germany following reunification: an empirical analysis using decomposable inequality measures, *Review of Income and Wealth*, 42, 1-11.
- Slottje, D., R.L.Basman and M.Nieswiadomy (1989), On the empirical relationship between several well-known inequality measures, *Journal of Econometrics*, 42,49-66.
- Vazquez, R., B. Melenberg and A. van Soest, (1999), Nonparametric bounds on the income

distribution in the presence of sample nonresponse, CentER DP paper 9933, *Center for Economic Research* (CentER), Tilburg University.

Vella, F., (1998), Estimating models with sample selection bias, *Journal of Human Resources* 33, 127-172.

Appendix A

This appendix show that the lower bound on the inter-quartile range given in (12) is sharp, but that the upper bound in (12) is not. Moreover, the upper bound in (15) is derived and shown to be sharp. For notational convenience, the conditional covariates X are not explicitly mentioned. The lower and upper bounds on the distribution function given in (8) are denoted by F_L and F_U , and the measures $P(\delta=0)$ and $P(\delta=1)$ by p_0 and p_1 , respectively. Throughout the appendix it is assumed that $p_0 < 0.5$, and for convenience, that the distribution of Y among respondents and in the population as a whole are continuous with invertible distribution functions, so that quantiles are uniquely determined.

Based on the short hand notation, the right and left and hands side of (8) can be re-written as,

$$\begin{aligned} P(Y \leq z | \delta=1)P(\delta=1) &= F_1(z)p_1 = F_L(z) \\ P(Y \leq z | \delta=1)P(\delta=1) + P(\delta=0) &= F_1(z)p_1 + p_0 = F_U(z) \end{aligned} \tag{A.1}$$

and the bounds on the distribution function are given by,

$$F_L(z) \leq F(z) \leq F_U(z) \tag{A.2}$$

Likewise, the notation for the bounds on the α -quantile of the distribution can also be simplified so that

$$\begin{aligned} q(\alpha) &= \inf \{z: F(z) \geq \alpha\} \\ U(\alpha) &= \inf \{z: F_L(z) \geq \alpha\} \\ L(\alpha) &= \inf \{z: F_U(z) \geq \alpha\} \end{aligned} \tag{A.3}$$

Thus expression (10) can also be written as,

$$L(\alpha) \leq q(\alpha) \leq U(\alpha) \tag{A.4}$$

The definitions given in (A.1) to (A.4) immediately imply the following relation between $U(\alpha)$ and $L(\alpha)$ which will be used at a later stage:

$$U(\alpha) = L(\alpha + p_0) \quad (\text{A.5})$$

The bounds given in (A.4) straightforwardly lead to the bounds on (12) on the inter-quartile range such that

$$L(0.75) - U(0.25) \leq IQR \leq U(0.75) - L(0.25) \quad (\text{A.6})$$

First, the lower bound in (12) is shown to be sharp. This means that, for given F_L and P_0 (which also then determines F_U and P_1), there always exists a distribution of Y in the population of non-respondents that makes IQR equal to the lower bound in (A.6). It is easy to see that any distribution among non-respondents with $P(Y \leq U(0.25) | \delta = 0) = 0$ and $P(Y \leq L(0.75) | \delta = 0) = 1$ will be sufficient, since these conditions make $F(z) = F_L(z)$ for $z \leq q(0.25)$ and $F(z) = F_U(z)$ for $z \geq q(0.75)$, thus $q(0.25) = U(0.25)$, $q(0.75) = L(0.75)$ and thus $IQR = L(0.75) - U(0.25)$. Such a distribution for non-respondents is possible as long as $U(0.25) = L(0.25 + p_0) < L(0.75)$, that is, under the regularity condition that $p_0 < 0.5$ and that the distribution among respondents is continuous.

The next step is to derive the upper bound on IQR given in (15). Since (as it will be shown later) this upper bound is, in general, smaller than the upper bound in (12), this result will imply that the upper bound in (12) is not sharp. First, note that the function $F - F_L$ is increasing. This is because, for any $a < b$,

$$\begin{aligned} F(a) - F_L(b) - [F(a) - F_L(a)] &= F(b) - F(a) - [F_L(b) - F_L(a)] = \\ P(a < Y \leq b | \delta = 1)p_1 + P(a < Y \leq b | \delta = 0)p_0 - P(a < Y \leq b | \delta = 1)p_1 &= \\ P(a < Y \leq b | \delta = 0)p_0 &\geq 0 \end{aligned} \quad (\text{A.7})$$

This, together with the assumption that F is invertible (so that F_L is invertible with inverse $U(\alpha)$), implies that,

$$0.5 = F[q(0.75)] - F[q(0.25)] \geq F_L[q(0.75)] - F_L[q(0.25)] \quad (\text{A.8})$$

so that,

$$q(0.75) \leq U[0.5 + F_L[q(0.25)]] \quad (\text{A.9})$$

and also,

$$IQR \leq U[0.5 + F_L[q(0.25)]] - q(0.25) \quad (\text{A.10})$$

Expression (A.10) is not a useful upper bound on the IQR, since $q(0.25)$ is not observed. But it is known that $L(0.25) \leq q(0.25) \leq U(0.25)$, thus,

$$IQR \leq \max\{U[0.5 + F_L(t)] - t; L(0.25) \leq t \leq U(0.25)\} \quad (\text{A.11})$$

An alternative expression for (A.11) is given by

$$IQR \leq \max\{U[0.5 + p] - U(p); F_L[L(0.25)] \leq p \leq 0.25\} \quad (\text{A.12})$$

The right hand side of (A.12) shows the upper bound in expression (15). Note that this is indeed at most as large and generally smaller than the upper bound in (12), since

$$\begin{aligned} & \max\{U[0.5 + p] - U(p); F_L[L(0.25)] \leq p \leq 0.25\} \leq \\ & \max\{U[0.5 + p]; F_L[L(0.25)] \leq p \leq 0.25\} - \\ & \min\{U[p]; F_L[F_L[L(0.25)]] \leq p \leq 0.25\} = \\ & U(0.75) - L(0.25) \end{aligned} \quad (\text{A.13})$$

The final step is to prove that the upper bound derived above is sharp, that is, to show that for a given F_L, p_0 there some distribution of Y among non-respondents such that the upper bound is attained. From the derivation of the upper bound given above, it is clear that this means that the distribution of Y among non-respondents must be such that the following two conditions are satisfied:

$$(1) \quad F[q(0.75)] - F[(0.25)] = F_L[q(0.75)] - F_L[q(0.25)]$$

and (A.14)

$$(2) \quad U[0.5 + F_L(q(0.25))] - Q(0.25) = \max\{U[0.5 + F_L(t)] - t; L(0.25) \leq t \leq U(0.25)\}$$

Let t^* be the (or a) value of t for which the maximum in the right hand side of the second equation in (A.14) is attained (which depends only on the given F_L, p_0 , not on the distribution of Y among non-respondents). Then the question is to find a distribution function $F_0(y) = P(Y \leq y | \delta = 0)$ such that

$$(1) \quad 0.5 = F_L[q(0.75)] - F_L[q(0.25)]$$

and (A.15)

$$(2) \quad q(0.25) = t^*$$

Condition (2, A.15) means that $P(Y \leq t^*) = F_L(t^*) + F_0(t^*)p_0 = 0.25$, so $F_0(t^*) = [0.25 - F_L(t^*)]/p_0$. Since $L(0.25) \leq t^* \leq U(0.25)$, it implies $[0.25 - p_0] \leq F_L(t^*) \leq 0.25$, so $[(0.25 - F_L(t^*))]/p_0 \in [0, 1]$. Thus it is possible to choose F_0 so that (2, A.15) is satisfied. Using condition (2, A.15), (1, A.15) can be rewritten as either $F_L[q(0.75)] = 0.5 + F_L[t^*]$, or as $q(0.75) = U[0.5 + F_L(t^*)]$. This means that,

$$\begin{aligned}
& F[Y \leq U(0.5 + F_L(t^*))] = 0.75 \\
\Rightarrow & F_L[U(0.5 + F_L(t^*))] + F_0[U(0.5 + F_L(t^*))]p_0 = 0.75 \\
\Rightarrow & 0.5 + F_L(t^*) + F_0[U(0.5 + F_L(t^*))]p_0 = 0.75 \\
\Rightarrow & F_0(U(0.5 + F_L(t^*))) = (0.25 - F_L(t^*)) / p_0 = F_0(t^*)
\end{aligned} \tag{A.16}$$

Thus, F_0 should be such that there is no probability mass between t^* and $U[0.5 + F_L(t^*)]$ which is larger than t^* . Thus it is possible to choose F_0 such that both conditions are satisfied and the upper bound in (15) is sharp.

Appendix B

*Table.B1: Correcting for CPI changes, CPI values per region,
Base=1997*

| Population | 1990 | 1997 |
|------------------------|---------------------|-------------|
| Unified Germany | 84.4 (-0.13495) | 100.00 |
| West Germany | 86.24 (-0.12096) | 100.00 |
| East Germany | 72.5 (-0.2569) | 100.00 |

Source: Federal Statistical Office Germany,
<http://www.destatis.de/indicators/e/vpi001aj.html>