Bank i Kredyt 41 (4), 2010, 45-72

BANK I KREDYT
BANK & CREDIT

www.bankikredyt.nbp.pl
www.bankandcredit.nbp.pl

Income and consumption inequality in Poland, 1998–2008

Michał Brzeziński*, Krzysztof Kostro#

Submitted: 24 August 2009. Accepted: 15 June 2010.

Abstract

This paper estimates a variety of inequality indices to study the evolution of income and consumption inequality in Poland between 1998 and 2008. We use robust methods to adjust for the impact of extremely large observations. We also conduct statistical tests on inequality changes using methods, which account for the complexity of the household sample design. All analyses are performed for the entire population, for rural and urban subpopulations, and for the three largest cities. The main result is that during 1998–2008 there was a statistically significant rise in economic inequalities in Poland, which depending on the inequality index, ranged from 8.7% to 19.6% in case of income distribution and from 6.5% to 12.3% in case of consumption distribution. Among the studied subpopulations, economic inequalities are both the highest and the fastest-growing in Warsaw, where consumption inequality as measured by the Gini index increased during the studied period by as much as almost 23%.

Keywords: income inequality, consumption inequality, Pareto model, robust estimation, statistical inference, Poland

JEL: D31, D63, P36

* University of Warsaw, Faculty of Economic Sciences; e-mail: mbrzezinski@wne.uw.edu.pl.
# University of Warsaw, Faculty of Economic Sciences; Polish Academy of Sciences, Institute of Economics.

1. Introduction

There is an extensive empirical literature on the evolution of economic inequality in Poland during the last two decades (see, e.g., Daras et al. 2006; Keane, Prasad 2002; 2006; Milanović 1999; Newell, Socha 2007; Szulc 2000; 2003). These studies deal with a number of important methodological issues in the measurement of inequality including the problems of data quality, the changes in the design of the household surveys and methods of sample selection, the choice of the equivalence scales, and others. However, the overwhelming majority of them delivers exclusively point estimates of various inequality indices computed on the basis of the sample data. This approach suffers from at least two important drawbacks. First, providing only point estimates of inequality measures does not allow for reliable statistical inference on differences in the values of inequality measures when one wants to compare, for example, values of inequality indices over time or across subpopulations. Since inequality indices are almost always estimated from the household sample data, gathered by the means of random sampling from the given population, they are subject to random variation in the sample, which has to be accounted for by statistical inference. Without estimation of standard errors, confidence intervals and conducting statistical tests, researchers have to make inequality comparisons informally using only point estimates, visual inspection of the trends in the estimates or their own *a priori* subjective beliefs. Conclusions derived by these methods can easily be misleading or simply wrong. Therefore, it is necessary to use formal statistical inference methods to arrive at valid conclusions about the statistical significance of observed differences in the computed values of inequality measures.

Secondly, it has been shown that inequality indices and the associated methods of distributional analysis like stochastic dominance tests based on Lorenz curves are very sensitive to the presence of extreme values in the sample distributional data (Cowell, Victoria-Feser 1996; 2007). Most of inequality indices are not robust to very large or very small data values in the sense that one single extreme observation can bring estimates of inequality measures to arbitrarily small (or large) values. In practice, extreme values are often found especially in the upper tails of distributions of incomes, assets or consumption expenditures. Such observations can appear in the data due to some form of contamination caused by measurement or coding errors, but they can also represent real very large (or very small) incomes or consumption expenditures, which have leverage on estimated aggregate inequality measures. However, even in the second case it seems undesirable that a single observation affects heavily values of indicators summarizing the overall distribution of relevant variable in the society. In order to overcome this problem, one should use robust statistical methods, which provide remedies against possibly misleading estimates of inequality indices and their standard errors.

In this paper, we estimate a number of inequality indices calculated for both income and consumption distributions in Poland, together with their estimated asymptotically correct standard errors. These estimates are robust to extreme income and consumption expenditure values. Inequality measures considered are the Gini coefficient, three members of the generalized entropy (GE) class, which are otherwise known as half the squared coefficient of variation, the Theil index, and the mean logarithmic deviation, two members of the Atkinson class of inequality indices, and two quantile income shares. We use a variety of inequality measures to ensure that the results do not depend on the choice of indices. We then test the statistical hypotheses concerning changes in the inequality indices over time.

The paper uses data for 1998–2008, taken from the Household Budget Survey (HBS) study conducted yearly by the Central Statistical Office (CSO).[1] All calculations use detailed information about the design of the HBS sample (i.e. information about weighting, clustering and stratification) throughout the period to provide correct estimates of the standard errors. All analyses are performed for the entire population, for rural and urban subpopulations, and for the three largest cities (Warsaw, Krakow and Lodz).

The reminder of the paper is organized as follows. Section 2 provides a short review of the existing literature on recent trends in inequality in Poland. Statistical methods of estimating standard errors of inequality indices as well as methods of robust estimation are presented in section 3. Section 4 introduces the HBS data, while section 5 reports and discusses empirical results. The last section concludes.

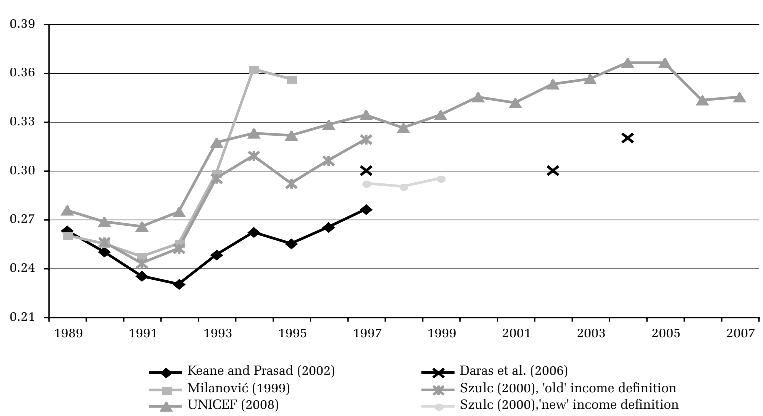
2. Review of the prior literature

Changes in economic inequality in Poland during economic transition have been analyzed in a number of studies. However, usually only a limited number of inequality indices were calculated with the Gini index being the most popular. For this reason, we discuss mainly results concerning this measure.

There is some disagreement about the trends in the Gini index for Poland, especially during the early phase of the transition (from 1992 to 1994). Figure 1 presents estimates of the Gini index for disposable income from five well-documented sources using HBS as the main data set. These studies are based on rather different methodologies. Keane and Prasad (2002) use individual data

Figure 1

Evolution of the Gini index for Poland according to various sources



[1] This is the longest period for which a consistent series of HBS data can be constructed without making uncertain assumptions (see also section 4).

and apply food-share based equivalence scale, Milanović (1999) uses both grouped and individual data and applies simple *per capita* scale, UNICEF (2009) uses grouped data and *per capita* scale, Daras et al. (2006) rely on individual data and OECD equivalence scale, while Szulc (2000; 2003) uses individual data and an empirically estimated scale. Szulc (2003) offers two independent estimates of the Gini index for 1997, since in that year an important change in the HBS definition of the 'disposable income' was introduced (see section 4).

All of these studies suggest that the Gini index for disposable income fell between 1989 and 1991. This was probably caused by a significant decline in household real incomes in the first two years of the transition (1990–1991) associated with reduced variability in income distribution. Estimates from Szulc (2000), and especially Milanović (1999), show large increases in inequality between 1992 and 1994 to the levels significantly higher than in 1989. On the other hand, estimates from Keane and Prasad (2002) imply that in 1992 inequality was still falling, and only for 1993–1994 there was a mild rise of inequality, but in 1994 the Gini index was still no higher than in 1989. The difference between Keane and Prasad's (2002) estimates and these from other studies is large and ranges from 0.05 to 0.10 for 1994. It arises because Keane and Prasad's (2002) study is the only one that attempts to account formally for important methodological changes in the sampling frame of the HBS (see section 4). In particular, their paper develops a technique for adjusting data for a change from quarterly to monthly household data collection, which was introduced in 1993.

For the years 1995–1997 all relevant studies give the same picture – income inequality as measured by the Gini index was steadily increasing. After 1997, UNICEF's (2009) estimates suggest an increasing trend in the Gini coefficient from 0.334 in 1997 to 0.366 in 2005. This view is consistent with the findings of Newell and Socha (2007), who have shown that during 1998–2004 wage inequality as measured by the Gini rose sharply from 0.231 to 0.262. To summarize, it seems that starting in 1992 income inequality grew substantially in Poland, and that this trend continued until the mid-2000s.

Several of the papers verify if results concerning inequality trends in Poland are robust to the choices of welfare measure (i.e. income vs. consumption) and inequality index. For example, Keane and Prasad (2002) apply the Gini index, the mean log deviation, the coefficient of variation and two quantile ratios to both income and consumption distributions. They find that inequality trends over the period 1989–1997 are rather insensitive to the choices of welfare measure and inequality index. Similar conclusions are drawn in Mitra and Yemtsov (2006) and Szulc (2003; 2008).

It is also interesting to put inequality estimates for Poland in international perspective. According to Brandolini and Smeeding (2008), the Gini value for the distribution of disposable income in Poland in 1999 was 0.29 – a number equal to the simple average of Ginis calculated for seventeen middle- and high-income economies.[2] Mitra and Yemtsov (2006) provide an informative comparison between Poland and other transition countries. In general, all Eastern European countries and post-Soviet states have experienced an increase in inequality during the process of transition. According to authors, since early 1990s consumption inequality in Poland as measured by the Gini index rose steadily, but gradually, reaching the value of 0.32 in 2002. Again, this value was very close to the simple average (0.318) of Ginis for the sample of eighteen transition countries considered in the paper.

[2] However, according to Brandolini and Smeeding (2008), the Gini index for market incomes (i.e. incomes before taxation and social transfers) is equal to 0.5 It is slightly lower than the corresponding numbers for the United Kingdom (0.51) and Israel (0.52), but higher than the Gini for the United States (0.47), Australia (0.46), France (0.49), Germany (0.47) or Canada (0.42). In other words, Polish economy generates a rather high level of income inequality, which is substantially reduced by the tax and transfer system.

3. Statistical methods

3.1. Estimation of inequality indices and their sampling variances from complex survey data

Although it was often implicitly assumed that in income analyses economists deal with very large samples where precision of the estimates is not problematic, Maasoumi (1997) observed that even for such samples the standard errors of inequality indices can be large. The case for computing variance estimates and performing statistical tests on the changes in values of inequality indices is even stronger if we take into account the fact that household surveys, which are a primary source of data for distributional analysis, are rarely simple random samples (SRSs), where every unit in the population has the same probability of being included in the sample. They are usually complex surveys with probability weighting of the units as well as clustering and stratification of the population (see, e.g., Kish, Frankel 1974).[3] Ignoring complexity of the survey design can lead to incorrect point estimates of population parameters (i.e. inequality indices) and inconsistent (usually underestimated) standard errors of these estimates.[4]

Recently, there have been significant developments in both theory and the practice of statistical inference on inequality indices estimated from survey data. In general, two types of inference have been developed using either approximate asymptotic or re-sampling based simulation methods. Asymptotically correct approximate standard errors for some inequality measures estimated from SRSs were derived, among others, by Cowell (1989) and Davidson (2009).[5] Other authors (Mills, Zandvakili 1997; Biewen 2002) proposed computationally intensive re-sampling methods such as the bootstrap for variance estimation of the most widely-used inequality indices in the SRS case. What is more interesting from the practically--oriented perspective, there have been are also a few studies delivering variance estimators for inequality measures calculated from the complex survey data. Binder and Kovačević (1995) and Kovačević and Binder (1997) proposed estimators accounting for the complex survey features for, among others, the Gini index, coefficient of variation and Lorenz curve ordinates. Biewen and Jenkins (2006) derived asymptotic expressions for sampling variances of the popular Atkinson and GE classes of inequality measures. Recently Bhattacharya (2007) has provided a general theory of asymptotic inference for Lorenz curves and the Gini coefficient with complex survey data. His formula is, however, much more difficult to implement in practice than these offered by earlier authors. Sampling variance of quantile share ratio in case of complex survey design was recently derived by Osier (2009) and Langel and Tillé (2009) using the generalized variance linearization method of Deville (1999).

Inequality measures considered in this paper are the Gini index, three members of the generalized entropy class (GE(0), GE(1), GE(2)), which are otherwise known respectively as the mean

---

[3] In a standard complex design for the household survey, prior to sampling the population is divided into a number of strata (e.g. administrative or geographical regions). Next, a sample of clusters (e.g. cities, counties, etc.) is drawn by simple random sampling with replacement from each stratum. Finally, a sample of households is drawn from each cluster.

[4] Estimation of standard errors accounting for the complexity of survey design in the context of poverty analysis for Poland is addressed in Szulc (2006).

[5] For short reviews of other asymptotic approaches see Biewen and Jenkins (2006, p. 372) and Cowell (2008, p. 186).

Table 1

Inequality indices and their asymptotic variances

Index Gini

Definition

$$G=\frac{1}{2N^2\mu}\sum\nolimits_{i=1}^{N}\sum\nolimits_{j=1}^{N}\left|y_i-y_j\right|$$

Asymptotic variance estimator term $\tilde{s}_{hij}$

$$\tilde{s}_{hij}=\frac{2}{\hat{\mu}}\left[\hat{A}(y_{hij})y_{hij}+\hat{B}(y_{hij})-\frac{\hat{\mu}}{2}\left(\hat{G}+1\right)\right]$$

where $\hat{A}(y)=\hat{F}(y)-\frac{\hat{G}+1}{2}$, $\hat{F}(y)$ is the empirical cumulative distribution function, $\hat{G}$ is sample estimate of G, $\hat{\mu}$ is the sample mean income and with $\hat{B}(y)=\sum\nolimits_{h=1}^{L}\sum\nolimits_{i=1}^{n_h}\sum\nolimits_{j=1}^{m_i}w_{hij}y_{hij}I(y_{hij}\geq y)$ with $I(\cdot)$ being an indicator function

Index GE(α)

Definition

$$GE(0)=\frac{1}{N}\sum\nolimits_{i=1}^{N}\ln\left(\frac{y_i}{\mu}\right)$$

Asymptotic variance estimator term $\tilde{s}_{hij}$

$$\tilde{s}_{hij}^{GE(0)}=-\hat{U}_0^{-1}\log y_{hij}+\hat{U}_1^{-1}y_{hij}+\hat{U}_0^{-1}(\hat{T}_0\hat{U}_0^{-1}-1)$$

Definition

$$GE(1)=\frac{1}{N}\sum\nolimits_{i=1}^{N}\left(\frac{y_i}{\mu}\right)\ln\left(\frac{y_i}{\mu}\right)$$

Asymptotic variance estimator term $\tilde{s}_{hij}$

$$\tilde{s}_{hij}^{GE(1)}=\hat{U}_1^{-1}y_{hij}\log y_{hij}-\hat{U}_1^{-1}(\hat{T}_1\hat{U}_1^{-1}+1)y_{hij}+\hat{U}_0^{-1}$$

Definition

$$GE(\alpha)=\frac{1}{\alpha^2-\alpha}\left[\frac{1}{N}\sum\nolimits_{i=1}^{N}\left(\begin{matrix}\frac{y_i}{\mu}\\ \mu\end{matrix}\right)^{\alpha}-1\right],\alpha\geq 2$$

Asymptotic variance estimator term $\tilde{s}_{hij}$

$$\tilde{s}_{hij}^{GE(\alpha)}=\frac{1}{\alpha}\hat{U}_\alpha\hat{U}_1^{-\alpha}\hat{U}_0^{\alpha-2}-\frac{1}{\alpha-1}\hat{U}_\alpha\hat{U}_1^{-\alpha-1}\hat{U}_0^{\alpha-1}y_{hij}+\frac{1}{\alpha^2-\alpha}U_0^{\alpha-1}U_1^{-\alpha}(y_{hij})^{\alpha}$$

where $\hat{U}_\alpha=\sum\nolimits_{h=1}^{L}\sum\nolimits_{i}^{n_h}\sum\nolimits_{j=1}^{m_i}w_{hij}(y_{hij})^{\alpha}$, $\hat{T}_\alpha=\sum\nolimits_{h=1}^{L}\sum\nolimits_{i}^{n_h}\sum\nolimits_{j=1}^{m_i}w_{hij}(y_{hij})^{\alpha}(\log y_{hij})$

Index A(ε)

Definition

$$A(\varepsilon)=1-\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\mu}\right)^{1-\varepsilon}\right]^{1/(1-\varepsilon)}$$

Asymptotic variance estimator term $\tilde{s}_{hij}$

$$\tilde{s}_{hij}^{A(\varepsilon)}=\frac{\varepsilon}{1-\varepsilon}U_1^{-1}U_{1-\varepsilon}^{1/(1-\varepsilon)}U_0^{-(1/1-\varepsilon)}+U_0^{-\varepsilon/(1-\varepsilon)}U_{1-\varepsilon}^{1/(1-\varepsilon)}U_1^{-2}y_{hij}-\frac{1}{1-\varepsilon}U_0^{-\varepsilon/(1-\varepsilon)}U_1^{-1}U_{1-\varepsilon}^{\varepsilon/(1-\varepsilon)}(y_{hij})^{1-\varepsilon}$$

Definition

$$A(1)=1-\frac{\prod_{i=1}^{N}\left(y_i^{1/N}\right)}{\mu}$$

Asymptotic variance estimator term $\tilde{s}_{hij}$

$$\tilde{s}_{hij}^{A(1)}=(\hat{A}(1)-1)\hat{U}_0^{-1}(1-\hat{U}_0^{-1}\hat{T}_0)+(1-\hat{A}(1)\hat{U}_1^{-1}y_{hij}+(\hat{A}(1)-1)\hat{U}_0^{-1}\log y_{hij}$$

where $\hat{U}$ and $\hat{T}$ are defined as for GE(α), and $\hat{A}$ is sample estimate of A

Index QSR

Definition

$$QSR=\frac{Y-Y_\alpha}{Y_{1-\alpha}}$$

Asymptotic variance estimator term $\tilde{s}_{hij}$

$$\hat{s}_{hij}^{QSR(\alpha)}=\frac{y_{hij}-IF(\hat{Y}_\alpha)_{hij}}{-\hat{Y}_{1-\alpha}}-\frac{(\hat{Y}-\hat{Y}_\alpha)IF(\hat{Y}_{1-\alpha})_{hij}}{\hat{Y}_{1-\alpha}^2}$$

where $IF(\hat{Y}_\tau)_{hij}=\tau\hat{Q}_\tau-(\hat{Q}_\tau-y_{hij})I(y_{hij}\leq\hat{Q}_\tau)$, $\hat{Q}_\alpha$ is the estimate of the quantile of order α, $\hat{Y}$ is the estimator of the total income and $\hat{Y}_\alpha$ is the sum of incomes up to $\hat{Q}_\alpha$ and $I(\cdot)$ again an indicator function

Note: y_i denotes income of individual *i*, w_i – sampling weight of individual *i*, *N* – the number of individuals in society, μ – mean income, *Y* – total income and Y_α – sum of incomes up to quantile of order α. See text for the interpretation of parameters α and ε for, respectively, GE and Atkinson indices.

logarithmic deviation (MLD), the Theil index and half the squared coefficient of variation, two members of the Atkinson class of inequality indices (A(0.5), A(2)), and two quantile share ratios (QSR(0.8), QSR(0.9)). Table 1 provides definitions of these inequality indices together with formulas for their asymptotic sampling variances.[6]

Inequality measures considered differ in their sensitivity to income differences in different parts of the distribution. For GE and Atkinson classes of indices, this sensitivity varies with the values of the parameters α and ε, respectively. In particular, the more positive the values of α, the more sensitive is the given GE index to differences in income shares among the top incomes. For the Atkinson class, larger values of inequality-aversion parameter ε correspond to a greater aversion to inequality differences among lower incomes. The Gini coefficient is most sensitive to income differences around the mode of the distribution. Quantile (e.g. quintile or decile) share ratios are rather poor inequality indices since they use only information about the top and the bottom quantile. However, they are popular measures as they are easily interpretable (e.g. QSR(0.8) is the ratio of the total income received by the richest 20% of the population to that received by the poorest 20%). Quintile share ratio – QSR(0.8) – along with the Gini index is a 'Laeken indicator' chosen by the European Union to officially monitor income inequality in EU Member States.

The formula for variance estimate of the Gini index in the complex survey data framework was taken from the works of Binder and Kovačević (1995) and Kovačević and Binder (1997). We also used results of Biewen and Jenkins (2006) in case of GE and Atkinson indices and of Langel and Tillé (2009) in case of quantile share ratios. Although these papers use more or less different approaches to derive sampling variances of inequality indices, all resulting formulae can be reduced to the well-known expression for the sampling variance of a total estimator (see, e.g., Deaton 1997). This result can be stated in a following general setup. Let the population be stratified into L strata (e.g. geographical or administrative regions) with N_h primary sampling units (PSUs) (e.g., cities, counties, etc.) in the h-th stratum, and M_i individuals in cluster i. In the first stage of sampling, $n_h(\geq 2)$ PSUs (clusters) are selected from stratum h, while in the second stage m_i last sampling units (LSUs) (e.g. households) are selected in PSU i, $i = 1,\ldots, n_h$. The variable of interest (e.g. household equivalent disposable income) is y_{hij} and w_{hii} is the sampling weight of LSU hij.[7] For any of the inequality indices (I) used in this paper, the variance estimate for its sample estimate $\hat{I}$ is given by

$$\hat{V}(\hat{I}) = \sum_{h=1}^{L} n_h/(n_h - 1) \sum_{i=1}^{n_h} \left(\sum_{j=1}^{m_i} w_{hij}\tilde{s}_{hij} - \frac{\sum_{i=1}^{n_h}\sum_{j=1}^{m_i} w_{hij}\tilde{s}_{hij}}{n_h} \right)^2 \quad (1)$$

with $\tilde{s}_{hij}$ defined for each survey estimate of inequality index as in Table 1.

Although the variance estimators given by equation (1) appear to be complicated, they can be relatively easily programmed in any statistical software offering commands for complex survey data analysis (see, e.g., StataCorp. 2009). In this paper, we use Stata programs *svylorenz*

[6] See Cowell (2000; 2008) and Jenkins and Van Kerm (2009) for a general exposition of these and other inequality measures.

[7] If the distribution of income among persons is analysed, then household sample weights have to be multiplied by corresponding household sizes.

developed by Jenkins (2006), and *svygei_svyatk* by Jenkins and Biewen (2005), which provide point estimates and sampling variances for all but one inequality index considered.[8]

In section 4, we use variance estimates calculated according to formulae from Table 1 to test for statistically significant changes in values of inequality indices between two distributions in different years. We use pairwise difference-in-means t-tests for independent samples and so the test statistic for a comparison between year A and year B is

$$W = \frac{\hat{I}_A - \hat{I}_B}{\left[\hat{V}(\hat{I}_A) - \hat{V}(\hat{I}_B)\right]^{1/2}} \tag{2}$$

where $\hat{I}_A$ and $\hat{I}_B$ are estimates of inequality indices in years A and B, while $\hat{V}(\hat{I}_A)$ and $\hat{V}(\hat{I}_B)$ are the corresponding estimates of variances. The null hypothesis of equality of inequality measures is rejected when P value calculated on the basis of Student's t distribution with $\sum_{h=1}^{L} n_h - L$ degrees of freedom is smaller than the conventional significance level of 0.05 or 0.01.

3.2. Extreme incomes and estimation of inequality measures

An important problem in estimation of inequality indices from survey data is that the estimates are very sensitive to extreme observations. Most of inequality measures are not robust to the data contamination in either of the tails of the distribution. In other words, the presence of single one extremely small (or large) observation can bring estimates of inequality indices to arbitrarily small (or large) values (see Cowell, Victoria-Feser 1996). In particular, Cowell and Flaichaire (2007) obtained following results. GE measures with $\alpha > 1$ are very sensitive to high values in the data, while GE indices with $\alpha < 1$, and Atkinson measures with $\varepsilon > 1$ are very sensitive to small incomes in the data. The Gini index is less sensitive to contamination in the upper tail than GE indices. GE measures are less sensitive to large observations for smaller values of α.

In order to overcome this problem, several methods of adjusting data have been proposed in the literature (see, e.g., van Kerm 2007). Two common but naive methods are trimming (i.e. removing a fixed percentage of the highest and/or lowest values) and winsorizing (i.e. replacing extreme values with the values of trimming thresholds). Both of these adjustment procedures suffer from a drawback that they lose all information contained in the tails of the distribution where extreme data are dropped or replaced with chosen limiting values. It is therefore sensible to use more sophisticated approaches, which rely on parametric modelling of the tails by the methods robust to data contamination. The most common method of this kind proceeds by fitting robustly the Pareto distribution model to the upper tail of the empirical distribution.[9] Robust parametric estimates of the upper tail can then be combined with empirical distribution function for the rest of the data

---

[8] Stata program computing quantile share ratio and its sampling variance according to the formula given in Table 1 introduced by Langel and Tillé (2009), has been developed by authors and can be downloaded from http://coin.wne.uw.edu.pl/mbrzezinski/software/.

[9] In principle, it is possible to apply similar methods to the analysis of the lower tail as well (see van Kerm 2007). There are, however, some serious additional difficulties, which are not addressed by existing approaches (cf. Cowell, Flaichaire 2007, pp. 1053–1054). However, for some other technical reasons, we apply a simple adjusting procedure for negative and zero incomes (see section 4).

to obtain semi-parametric distribution for which standard distributional analyses can be applied (Cowell, Victoria-Feser 2007).

In this paper, we estimate upper tails of income and consumption distributions by fitting the Pareto distribution model with cumulative distribution function given by

$$F_{\theta,x_0}(x) = 1 - \left[\frac{x}{x_0}\right]^{-\theta}, \; x \geq x_0 \tag{3}$$

where θ is a shape parameter known as the Pareto index, and x_0 is a quantile above which Pareto model is assumed to be a correct one. Pareto model is usually estimated by maximum likelihood estimator (MLE). However, the MLE for the Pareto model is not robust to extreme observations (Victoria-Feser, Ronchetti 1994). Specifically for income distribution models, Victoria-Feser and Ronchetti (1994) proposed robust estimators known as optimal B-robust estimators (OBRE) (see also Cowell, Victoria-Feser 1996; 2007). For a sample of n observations and a given robustness constant c, OBRE is defined as a solution of the system of equations

$$\sum\nolimits_{i=1}^{n} \psi(x_i;\, \theta) = \sum\nolimits_{i=1}^{n} \left\{ s(x_i;\, \theta) - a(\theta) \right\} W_c(x_i; \theta) = 0 \tag{4}$$

with

$$W_c(x;\, \theta) = \min \left\{ 1;\; \frac{c}{\left\| A(\theta)\left[s(x;\theta) - a(\theta) \right] \right\|} \right\} \tag{5}$$

where ||·|| denotes the Euclidean norm, and the matrix $A(\theta)$ and vector $a(\theta)$ are defined implicitly by

$$E\left[\psi(x_i;\theta)\,\psi(x_i;\theta)^T\right] = \left[A(\theta)^T A(\theta)\right]^{-1} \tag{6}$$

$$E\left[\psi(x_i;\, \theta)\right] = 0. \tag{7}$$

For the Pareto model, the score function $s(x;\, \theta)$ is given by: $s(x;\theta) = \frac{1}{\theta} - \log(x) + \log(x_0)$. The robustness weights given in equation (5) are attributed to each observation and show how much an observation deviates from the assumed model. The values of these weights fall between 0 and 1. An observation is consistent with the Pareto model if its weight is equal to 1, however if a weight if less than 1 an observation should be downweighted as it is an outlier for the model. These weights can be then used to adjust the upper tail of the empirical distribution before standard procedures for estimating inequality indices can be applied. In section 5, we use this approach by multiplying sampling weights by robustness weights W_c defined in equation (5) (cf. Van Kerm 2007, p. 9).

In computation of OBRE and associated robustness weights we use iterative stepwise algorithm proposed by Victoria-Feser and Ronchetti (1997), which updates in turn $A(\theta)$, $a(\theta)$ and θ. We follow Cowell and Victoria-Feser (2007) in setting robustness constant c to 2, which for the Pareto model leads to an OBRE achieving approximately 85% efficiency in comparison with MLE. Finally, before computing OBRE, we use the prediction error criterion (C-criterion) proposed by Dupuis and Victoria-Feser (2006) for estimation of parameter x_0 for the Pareto model.[10]

[10] Stata programs implementing algorithms for computing C-criterion and OBRE for the Pareto model can be downloaded from the webpage http://coin.wne.uw.edu.pl/mbrzezinski/software/.

4. Data

We use yearly HBS micro-data for the period 1998–2008.[11] Before 1993, the survey did not cover properly several groups such as self-employed outside agriculture, social welfare recipients, as well as security, police and military personnel. In 1993 two major changes were introduced. First, HBS became fully representative for all main socio-economic types of households. Second, a new method of rotating households was applied – monthly rotation replaced previously used quarterly rotation. For these reasons, pre-1993 and post-1993 HBS data are not directly comparable unless some adjustment procedure is applied (cf. Keane, Prasad 2002). Another important modification of the HBS occurred experimentally in 1997, and definitively in 1998, when in order to adjust HBS to Eurostat recommendations, new definitions of some core concepts (i.e. disposable income) were implemented. Again, due to this change it is rather difficult to construct fully comparable data series for the period before 1998 and after this year.[12] Therefore, in this paper we use HBS data from 1998 to 2008 (the last available year).

HBS uses a two-stage stratified sampling scheme. In the first stage, the population is divided into a fixed number of strata from which primary sampling units (PSUs), that is clusters, are randomly chosen. PSUs consist of enumeration statistical districts (ESDs) or clusters of ESDs covering at least 250 dwellings. In the second stage of sample selection, dwellings are randomly selected from the PSUs selected in the first stage. Sample sizes are rather large and range from 31 428 to 37 366 households. In 2000 and 2001 there were several changes in the design of the stratification of population (Kordos et al. 2002, pp. 565–567). From 2001 on, the population has been stratified in 96 strata by voivodships, and in each voivodship by the size of the cities or in case of rural areas by groups of counties (*powiats*). In section 5, we use the detailed information about stratification and clustering of the HBS samples for every year to calculate corrected variance estimates for inequality measures.

Household net disposable income (i.e. post-tax-and-transfer income) is the main income concept used. It includes cash wages and salaries, self-employment income (including farm income), cash property income, social transfers (including social insurance, social assistance) and other income. Income taxes, mandatory payroll taxes and gifts donated to other households are not included. As a consumption measure we use total expenditures on consumer goods and services, which include expenditures on food, clothing, housing, health care, transportation and communication, culture and recreation and education. It includes expenses on durables and natural consumption.

An important methodological problem in estimating inequality indices from survey data that of negative and zero incomes. Since many standard inequality measures are undefined or are not 'well-behaved' indices in presence of negative and zero values (cf. Amiel et al. 1996), we replace such incomes with household's consumption expenditures.

We consider the individual as the main unit of analysis. In order to obtain personal distributions, all household observations are weighted by the product of household weights provided by the HBS

---

[11] The detailed description of the HBS design and its other features can be found in Kordos et al. (2002) and Central Statistical Office (2008).

[12] Various other shortcomings of the HBS data are nicely summarized in Levy, Morawski (2007).

Figure 2

Income and consumption inequality in Poland, 1998–2008

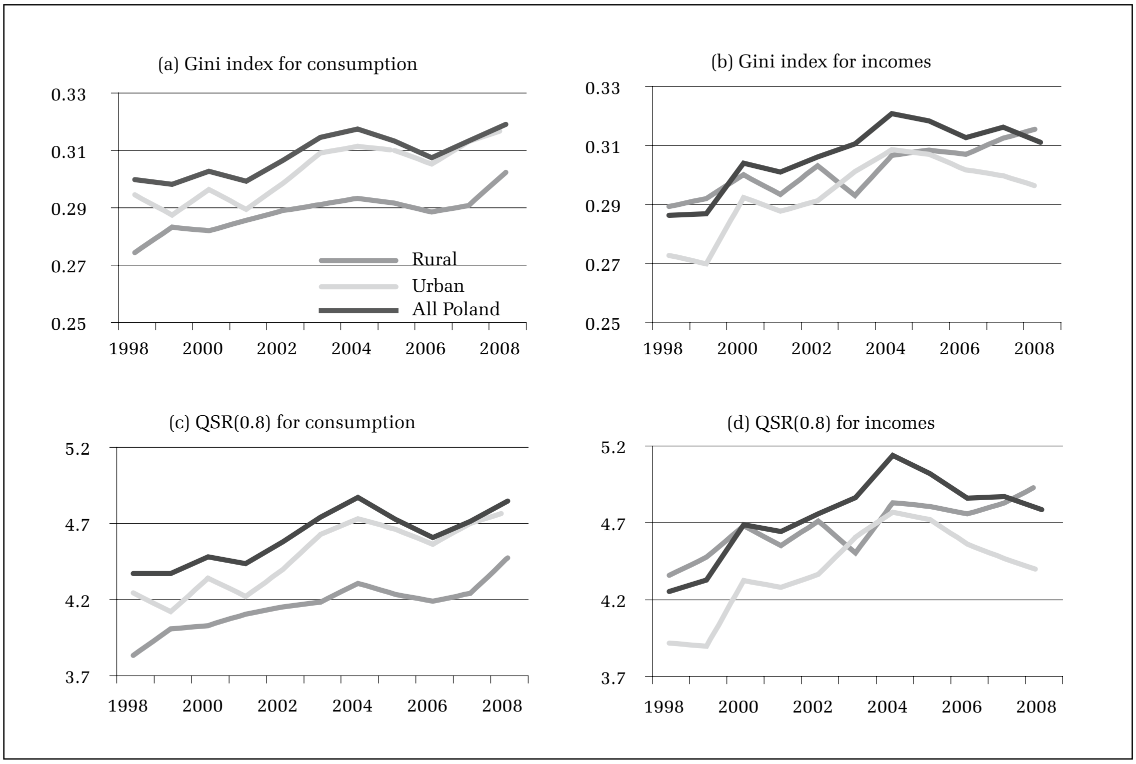
Table 2

Inequality estimates and tests for changes in inequality indices (all Poland)

Index	**Year**			**Changes in inequality indices (%) (*P* values)**		
	1998	**2003**	**2008**	**2008 vs. 1998**	**2008 vs. 2003**	**2003 vs. 1998**
			Disposable incomes			
Gini	0.286	0.310	0.310	8.7	0.1	8.5
	(0.002)	(0.002)	(0.003)	(0.000)	(0.899)	(0.000)
MLD	0.143	0.167	0.168	17.4	0.3	17.0
	(0.002)	(0.003)	(0.003)	(0.000)	(0.894)	(0.000)
Theil	0.151	0.176	0.181	19.6	2.8	16.2
	(0.004)	(0.005)	(0.005)	(0.000)	(0.447)	(0.000)
GE(2)	0.218	0.260	0.308	40.8	18.2	19.1
	(0.012)	(0.017)	(0.033)	(0.011)	(0.200)	(0.045)
Atkinson(0.5)	0.070	0.081	0.082	17.5	1.2	16.1
	(0.001)	(0.002)	(0.002)	(0.000)	(0.671)	(0.000)
Atkinson(2)	0.333	0.329	0.352	5.6	7.1	-1.3
	(0.040)	(0.011)	(0.019)	(0.669)	(0.282)	(0.914)
QSR(0.8)	4.239	4.850	4.774	12.6	-1.6	14.4
	(0.044)	(0.056)	(0.055)	(0.000)	(0.332)	(0.000)
QSR(0.9)	6.576	7.751	7.643	16.2	-1.4	17.9
	(0.099)	(0.128)	(0.127)	(0.000)	(0.549)	(0.000)
			Consumption expenditures			
Gini	0.299	0.314	0.318	6.5	1.5	4.9
	(0.002)	(0.002)	(0.002)	(0.000)	(0.143)	(0.000)
MLD	0.149	0.162	0.167	12.3	2.9	9.0
	(0.003)	(0.002)	(0.002)	(0.000)	(0.159)	(0.000)
Theil	0.174	0.183	0.189	8.3	3.4	4.8
	(0.004)	(0.004)	(0.004)	(0.012)	(0.247)	(0.143)
GE(2)	0.285	0.275	0.290	1.8	5.6	-3.6
	(0.014)	(0.011)	(0.012)	(0.780)	(0.337)	(0.555)
Atkinson(0.5)	0.077	0.082	0.084	10.2	3.0	7.0
	(0.001)	(0.001)	(0.001)	(0.000)	(0.197)	(0.007)
Atkinson(2)	0.238	0.260	0.266	11.8	2.3	9.3
	(0.003)	(0.003)	(0.003)	(0.000)	(0.120)	(0.000)
QSR(0.8)	4.358	4.729	4.835	10.9	2.2	8.5
	(0.047)	(0.047)	(0.049)	(0.000)	(0.117)	(0.000)
QSR(0.9)	6.575	7.182	7.372	12.1	2.6	9.2
	(0.097)	(0.098)	(0.102)	(0.000)	(0.180)	(0.000)

Notes: Columns from two to four give estimates of inequality indices and their standard errors (in parentheses). The last three columns report, first, the percentage change in the inequality index and, second, *P* value from the test for the equality of the inequality indices in respective years (see section 3.2).

Figure 3
The Gini and QSR(0.8) indices for rural and urban subpopulations, 1998–2008



and household sizes.[13] We use CPI deflators provided by CSO to adjust for differences in the prices faced by households in different years and/or regions. For income distributions, we use monthly price indices of consumer goods and services specific for five socio-economic groups. In case of consumption expenditures, we have used quarterly consumer price indices for voivodships and 12 categories of consumption expenditures. All distributions have been expressed in December 2008 price levels.[14] Finally, in order to adjust for the size and composition of households, all incomes are divided by the original OECD equivalence scale, which assigns weights 0.7 to any adult household member beyond the first and 0.5 to children under 14 years old.[15]

[13] HBS weights are non-response weights adjusting sample data for the differential non-response rates of different types of households. The method of estimating these weights has changed several times between 1998 and 2004. See Kordos et al. (2002) and Central Statistical Office (2008) for details. As indicated earlier, HBS weights are weighted also by OBRE robustness weights (see section 3.2).

[14] Due to data limitations, we have used linear interpolation to achieve monthly deflators for consumption distributions. We have also assumed that the structure of regional prices during 1998–1999 matches that of 2000.

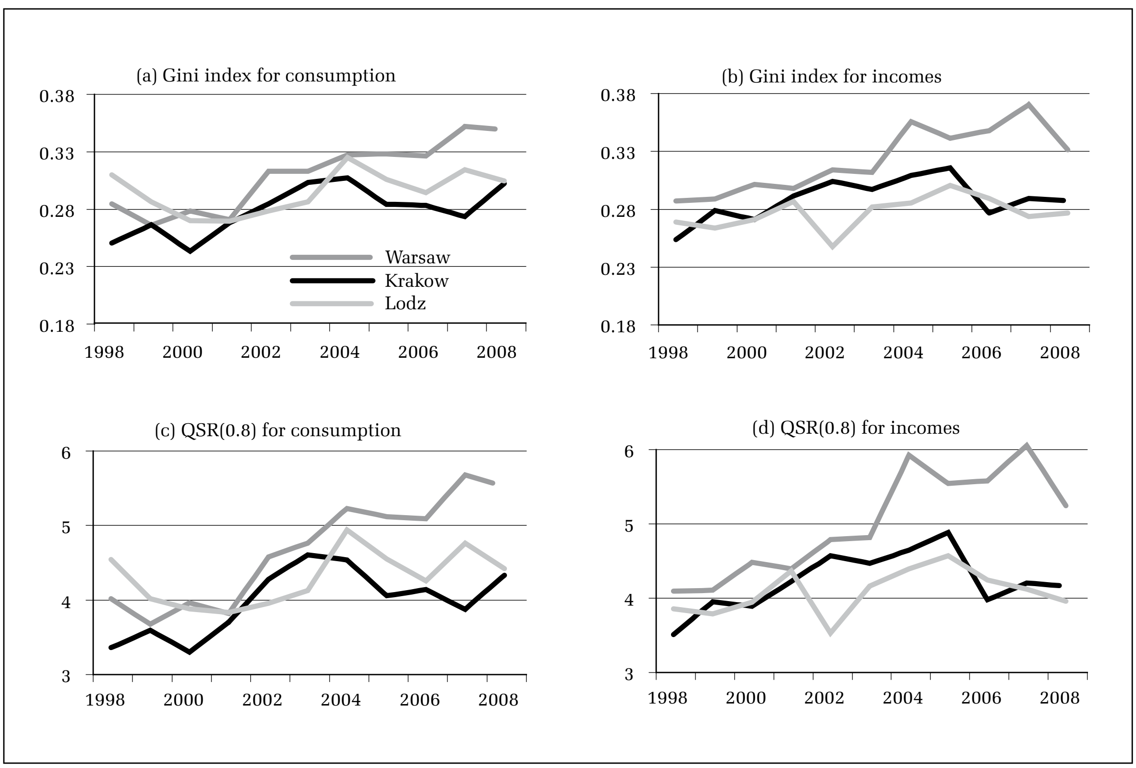
[15] Szulc (2006) argued convincingly that for Poland the original OECD scale is more appropriate than the modified OECD scale and other non-estimated scales as economies of scale in Polish households are rather low due to the relatively high expenditures on food and relatively low expenditures on housing. This point is also discussed in Brandolini (2007).

5. Results and discussion

In our estimations, we have fitted the Pareto model to the upper tail of the distributions using OBRE presented in section 3.2. We have also applied all other data adjustment procedures described in section 4. Tables 1–6 in the Appendix report changes in disposable incomes and consumption expenditures according to the mean value and various percentiles for the entire population and for the five subpopulations studied. In case of all Poland, both mean and median incomes as well as incomes at the first and the tenth percentiles displayed a slowly declining trend during 1998–2004. All of these statistics have been rising since 2004 with a big acceleration from 2006 onwards. On the other hand, incomes at the 90th and the 99th percentiles remained initially roughly stable until, respectively, 2005 and 2002, and then started to increase with a big acceleration in the last three years. All analyzed income statistics have higher levels in 2008 than in 1998 with the most gain for incomes at the 90th percentile (28.2%) and the 99th percentile (35.0%). The distribution of consumption expenditures experienced similar changes, but it is less volatile than income distribution.

Figure 2 shows the evolution of our eight inequality indices calculated for all Poland and for both data series used. For majority of indices, consumption and income inequalities display a similar level and follow a similar trend. Poland has experienced a rather slow but steady rise in both income and consumption inequality during 1998–2004. After that, inequality dropped a little

Figure 4
The Gini and QSR(0.8) indices for three major cities, 1998–2008



between 2004 and 2006, but in the last two years under study started to increase again. In overall, inequality levels are higher in 2008 than in 1998 for all inequality measures and for both welfare indicators used.

Two interesting results deserve further comments. First, in case of income distribution, despite using robust estimation methods, GE(2) index, which is very sensitive to extreme observations in the upper tail, still behaves in a clearly more unstable way than other measures (see Figure 2,

Table 3

Inequality estimates and tests for changes in inequality indices (rural Poland)

Index	**Year**			**Changes in inequality indices (%) (*P* values)**		
	1998	**2003**	**2008**	**2008 vs. 1998**	**2008 vs. 2003**	**2003 vs. 1998**
			Disposable incomes			
Gini	0.289	0.292	0.315	9.2	7.9	1.3
	(0.005)	(0.004)	(0.004)	(0.000)	(0.000)	(0.529)
MLD	0.152	0.153	0.181	19.5	18.1	1.2
	(0.005)	(0.004)	(0.006)	(0.000)	(0.000)	(0.780)
Theil	0.161	0.155	0.204	26.7	31.3	-3.5
	(0.009)	(0.005)	(0.012)	(0.004)	(0.000)	(0.576)
GE(2)	0.252	0.214	0.489	94.4	128.8	-15.0
	(0.031)	(0.015)	(0.129)	(0.073)	(0.034)	(0.272)
Atkinson(0.5)	0.074	0.073	0.089	20.7	21.4	-0.6
	(0.003)	(0.002)	(0.003)	(0.001)	(0.000)	(0.907)
Atkinson(2)	0.329	0.325	0.416	26.4	27.9	-1.2
	(0.022)	(0.015)	(0.032)	(0.026)	(0.011)	(0.886)
QSR(0.8)	4.344	4.492	4.942	13.8	10.0	3.4
	(0.088)	(0.079)	(0.099)	(0.000)	(0.000)	(0.209)
QSR(0.9)	6.994	7.189	8.344	19.3	16.1	2.8
	(0.213)	(0.186)	(0.250)	(0.000)	(0.000)	(0.491)
			Consumption expenditures			
Gini	0.274	0.290	0.302	10.3	3.9	6.2
	(0.004)	(0.004)	(0.003)	(0.000)	(0.019)	(0.001)
MLD	0.123	0.138	0.150	22.2	8.3	12.8
	(0.003)	(0.004)	(0.003)	(0.000)	(0.018)	(0.002)
Theil	0.139	0.155	0.167	19.9	7.2	11.8
	(0.005)	(0.006)	(0.005)	(0.000)	(0.130)	(0.039)
GE(2)	0.202	0.226	0.239	18.0	5.8	11.5
	(0.015)	(0.017)	(0.011)	(0.054)	(0.522)	(0.303)
Atkinson(0.5)	0.063	0.070	0.075	20.4	7.5	12.0
	(0.002)	(0.002)	(0.002)	(0.000)	(0.051)	(0.009)
Atkinson(2)	0.202	0.226	0.244	20.7	7.9	11.9
	(0.004)	(0.004)	(0.004)	(0.000)	(0.002)	(0.000)
QSR(0.8)	3.821	4.169	4.459	16.7	6.9	9.1
	(0.059)	(0.068)	(0.065)	(0.000)	(0.002)	(0.000)
QSR(0.9)	5.504	6.135	6.681	21.4	8.9	11.5
	(0.119)	(0.139)	(0.131)	(0.000)	(0.004)	(0.001)

Notes: see Table 2.

panel d).[16] To analyze this in more detail, we focus on a big spike in GE(2) estimates between 2006 and 2007. This spike reflects the fact that in 2007 there are two very large incomes in the sample, both roughly two times larger than the third-largest income, while 2006 sample does not contain extreme incomes. However, if estimation is based solely on sample data without robust estimation

Table 4

Inequality estimates and tests for changes in inequality indices (urban Poland)

Index	**Year**			**Changes in inequality indices (%) (*P* values)**		
	1998	**2003**	**2008**	**2008 vs. 1998**	**2008 vs. 2003**	**2003 vs. 1998**
			Disposable incomes			
Gini	0.272	0.300	0.295	8.5	-1.8	10.5
	(0.002)	(0.003)	(0.003)	(0.000)	(0.205)	(0.000)
MLD	0.126	0.154	0.146	16.5	-5.2	22.9
	(0.003)	(0.003)	(0.003)	(0.000)	(0.071)	(0.000)
Theil	0.137	0.164	0.159	15.9	-3.5	20.1
	(0.004)	(0.005)	(0.005)	(0.000)	(0.393)	(0.000)
GE(2)	0.196	0.237	0.229	17.1	-3.1	20.8
	(0.013)	(0.016)	(0.015)	(0.093)	(0.736)	(0.046)
Atkinson(0.5)	0.063	0.076	0.073	15.9	-4.0	20.7
	(0.001)	(0.002)	(0.002)	(0.000)	(0.211)	(0.000)
Atkinson(2)	0.318	0.298	0.265	-16.8	-11.1	-6.4
	(0.072)	(0.016)	(0.010)	(0.463)	(0.078)	(0.784)
QSR(0.8)	3.904	4.593	4.372	12.0	-4.8	17.7
	(0.045)	(0.062)	(0.060)	(0.000)	(0.011)	(0.000)
QSR(0.9)	5.805	7.156	6.643	14.4	-7.2	23.3
	(0.096)	(0.137)	(0.128)	(0.000)	(0.006)	(0.000)
			Consumption expenditures			
Gini	0.294	0.308	0.317	7.8	2.7	4.9
	(0.003)	(0.003)	(0.003)	(0.000)	(0.034)	(0.000)
MLD	0.145	0.157	0.165	13.9	5.0	8.5
	(0.003)	(0.003)	(0.003)	(0.000)	(0.068)	(0.004)
Theil	0.173	0.177	0.189	9.5	6.8	2.6
	(0.005)	(0.004)	(0.005)	(0.029)	(0.077)	(0.520)
GE(2)	0.287	0.263	0.298	3.9	13.2	-8.3
	(0.016)	(0.011)	(0.016)	(0.631)	(0.082)	(0.237)
Atkinson(0.5)	0.075	0.080	0.084	11.6	5.5	5.8
	(0.002)	(0.002)	(0.002)	(0.001)	(0.069)	(0.074)
Atkinson(2)	0.232	0.255	0.262	12.9	3.1	9.5
	(0.004)	(0.003)	(0.004)	(0.000)	(0.114)	(0.000)
QSR(0.8)	4.232	4.614	4.766	12.6	3.3	9.0
	(0.058)	(0.057)	(0.062)	(0.000)	(0.071)	(0.000)
QSR(0.9)	6.463	7.056	7.235	11.9	2.5	9.2
	(0.125)	(0.118)	(0.130)	(0.000)	(0.309)	(0.001)

Notes: see Table 2.

[16] In principle it is possible to use more robust approach by setting higher robustness constant for the OBRE (see section 3.2). This comes, however, at a serious cost of lower efficiency for the estimator.

of the right tail, GE(2) estimate increases even more to 0.387, which is almost 10% higher than our robust estimate (0.353). Moreover, non-robust estimation increases the magnitude of GE(2) standard error by nearly 60%. It is therefore likely that statistical inference based on non-robust estimates would be highly unreliable.

Second, for Atkinson(2) and QSR(0.9) indices the level of income inequality is often significantly higher than the level of consumption inequality. In case of the former index, it suggests that there is more inequality in the lower tail of Polish income distributions compared with distributions of

Table 5

Inequality estimates and tests for changes in inequality indices (Warsaw, Krakow, Lodz)

Index	**Year**			**Changes in inequality indices (%) (*P* values)**		
	1998	**2003**	**2008**	**2008 vs. 1998**	**2008 vs. 2003**	**2003 vs. 1998**
			Disposable incomes			
Warsaw						
Gini	0.286	0.310	0.330	15.4	6.3	8.6
	(0.009)	(0.009)	(0.009)	(0.001)	(0.144)	(0.060)
QSR(0.8)	4.067	4.787	5.220	28.3	9.0	17.7
	(0.163)	(0.228)	(0.254)	(0.000)	(0.205)	(0.010)
Krakow						
Gini	0.252	0.296	0.285	13.2	-3.4	17.2
	(0.010)	(0.014)	(0.012)	(0.033)	(0.579)	(0.010)
QSR(0.8)	3.483	4.443	4.137	18.8	-6.9	27.6
	(0.179)	(0.305)	(0.251)	(0.034)	(0.439)	(0.007)
Lodz						
Gini	0.267	0.280	0.275	2.8	-2.0	4.9
	(0.006)	(0.010)	(0.011)	(0.561)	(0.712)	(0.279)
QSR(0.8)	3.827	4.135	3.935	2.8	-4.8	8.0
	(0.129)	(0.223)	(0.199)	(0.649)	(0.504)	(0.233)
			Consumption expenditures			
Warsaw						
Gini	0.283	0.311	0.347	22.8	11.6	10.0
	(0.009)	(0.009)	(0.011)	(0.000)	(0.012)	(0.027)
QSR(0.8)	3.992	4.734	5.493	37.6	16.0	18.6
	(0.164)	(0.203)	(0.260)	(0.000)	(0.022)	(0.005)
Krakow						
Gini	0.249	0.301	0.301	20.8	-0.3	21.1
	(0.009)	(0.010)	(0.011)	(0.000)	(0.959)	(0.000)
QSR(0.8)	3.332	4.577	4.304	29.2	-6.0	37.4
	(0.148)	(0.256)	(0.219)	(0.000)	(0.417)	(0.000)
Lodz						
Gini	0.308	0.285	0.303	-1.7	6.4	-7.6
	(0.013)	(0.011)	(0.014)	(0.782)	(0.304)	(0.163)
QSR(0.8)	4.514	4.102	4.397	-2.6	7.2	-9.1
	(0.259)	(0.203)	(0.276)	(0.757)	(0.390)	(0.211)

Notes: see Table 2.

consumption expenditures. In case of the latter one, it reflects the fact that the share of the total income received by the poorest 10% is in Poland usually significantly lower than the analogous share of consumption expenditure distributions.

Table 2 reports point estimates of inequality indices for the beginning, middle and end of the period under study (1998, 2003 and 2008) as well as results of statistical tests for pairwise comparisons between the three years. The increase in income inequality over the whole period studied ranges from 5.6 (Atkinson(2) index) to 40.8% (GE(2) index). It is the largest for top-sensitive measure GE(2), suggesting that increases in the number and the magnitude of top incomes are most responsible for the recent increase in income inequality in Poland.[17] Inequality of consumption expenditures has increased in the range from 1.8 (GE(2) index) to 12.3% (MLD). Increases in consumption inequality during 1998–2008 are smaller than increases in income inequality for all indices with the exception of one (bottom-sensitive) Atkinson(2) index. Table 2 confirms conclusions derived from graphical analysis that according to almost all of our indices the bulk of the rise in income and consumption inequalities occurred in the first half of the period. It is also instructive to compare our results for the Gini index with these of Figure 1. Our estimates are very similar to findings of Szulc (2003) and Daras et al. (2006), while the estimates from UNICEF (2009) are slightly higher but show the same trend.

The results of statistical tests in Table 2 suggest that for 0.05 significance level and for almost all measures used there was a statistically significant increase in both income and consumption inequalities during 1998–2008. A small rise in consumption inequality according to GE(2) index as well as in income inequality according to Atkinson(2) index are statistically insignificant. Moreover, results for changes in GE(2) index for income distribution and for changes in the Theil index for consumption distribution are only borderline significant at 0.01 level (P value equal to 0.012 and 0.011, respectively). We also find that for the subperiod 2003–2008 we cannot reject any of the null hypotheses of no change in inequality indices. On the other hand, changes in inequality indices for 1998–2003 subperiod are always statistically significant if inequality changes according to these indices are significant for the whole period 1998–2008.

Figure 3 shows the evolution of two inequality indices (the Gini index and QSR(0.8)) that belong to the 'Laeken indicators' group for rural and urban populations. For the ease of interpretation, Figure 3 plots results for all Poland as well. As far as inequality levels are concerned, consumption inequality is lower for rural population than for urban population. The opposite seems to be true for income inequality – for majority of years rural income distribution is more unequal than urban one. Both income and consumption inequalities have increased for rural and urban subpopulations.

Figure 4 extends this analysis to cover populations of the three largest Polish cities (Warsaw, Krakow and Lodz).[18] Visual inspection suggests that during 1998–2008 income and consumption inequalities increased for Warsaw and Krakow, but possibly did not change for Lodz. It seems also

[17] Tables 1–6 in the Appendix suggest that for all populations analyzed incomes and consumption expenditures at high percentiles (90th and 99th) usually grew much faster than at low and middle percentiles (1st, 10th, 50th).

[18] We have chosen these subpopulations mainly for illustrative purposes to show the effect of sample size on the results of statistical inference. The number of sample observation for the populations of the three cities is significantly smaller than the total HBS sample size and ranges from about 600 to 1800. However, it is large enough to allow for meaningful statistical inference. We leave the problem of analyzing inequalities across other subgroups (e.g. defined along region of residence, employment status, age or occupation) for future research.

that in general Warsaw have experienced a higher rise in inequalities than Krakow. Graphical analysis is confirmed by point estimates of inequality indices and results of statistical tests reported in Tables 3–5.

Table 3 shows that over the period under study there has been a sizable increase in income and consumption inequalities in rural Poland. Income inequality has risen in the range from 9.2 (the Gini index) to 26.7% (the Theil index), while consumption inequality in the range from 10.3 (the Gini index) to 22.2% (MLD).[19] The bulk of the increases in income inequality occurred after 2003; changes before 2003 are not statistically significant. On the other hand, consumption inequality has increased more during 1998–2003.

Results for the urban population are reported in Table 4. Economic inequalities in urban Poland have increased during 1998–2008, but at a slower pace than rural inequalities. Income inequality has increased from 8.5 (the Gini index) to 16.5% (MLD), while consumption inequality from 7.8 (the Gini index) to 13.9% (MLD). For most of the indices we cannot reject the null hypothesis of no change in inequality between 2003 and 2008. The only exceptions are quantile share ratios estimated for income distribution, which display moderate (from 4.8 to 7.2%) marginally significant drops during this subperiod.

Table 5 provides results for the two 'Laeken' inequality measures estimated for populations of Warsaw, Krakow and Lodz. Warsaw is clearly the most unequal among the largest Polish cities. For 2008, the Gini index for income inequality in Warsaw is 6.5% higher than the index for national inequality, and 11.9% higher than the Gini for overall urban population. The numbers for consumption inequality are, respectively, 9.1% and 9.5%. What is more striking, the Gini index for consumption inequality in Warsaw increased substantially during the entire period under study by as much as 22.8%. Income inequality as measured by the Gini rose by 15.4%. Even greater changes in inequalities are found if one uses QSR(0.8). According to this index, the ratio of the total income of the richest fifth to the total income of the poorest fifth has increased in Warsaw during 1998–2008 by 28.3%. The corresponding number for the distribution of consumption expenditures is even greater – 37.6%.

Estimates for Krakow suggest that inequalities grew there almost as fast as in Warsaw. However, only in case of consumption inequality, these changes are statistically significant at the 0.01 level. Lodz emerges as a city with a relatively stable income and consumption distributions. For every pairwise comparison among the three analyzed years, we cannot reject any of the null hypotheses of equality of the Ginis or QSR(0.8) indices for Lodz.

6. Conclusions

In this paper, we have used micro data from the Polish Household Budget Survey (HBS) to study the evolution of economic inequalities over the period 1998–2008. Our results are fairly robust to the choice of inequality indices, welfare indicators and the presence of extreme values in the upper tail of distributions. We have provided point estimates of inequality measures and also estimated sampling variances of inequality indices using methods that take into account full complexity of the sample design. This allowed us to conduct statistical tests verifying if observed inequality changes are statistically significant.

[19] From now on, we report only results significant at the 0.01 level.

Our major findings are the following. First, there was a rather slow but steady growth in both income and consumption inequalities, especially during 1998–2003. The exact magnitude of the increase depends on the inequality measure and welfare indicator used. For income distributions, the increase ranges from 8.7 to 19.6%, while for consumption expenditures it is a little smaller and ranges from 6.5 to 12.3%. Economic inequalities in rural Poland have been rising faster than inequalities in urban Poland. Analysis for the three major Polish cities suggests that Warsaw is the most unequal among them. The Gini index for consumption expenditures in Warsaw has grown by as much as about 23%.

The more general conclusions concern the evaluation of our methodological framework. First, we have shown that even in large samples (the full HBS sample), the use of non-robust methods can lead to large variability in point estimates and variance estimates for top-sensitive inequality indices (i.e. GE(2) index). This can increase the risk of misleading inferences about inequality trends and statistical significance of inequality changes. Second, our analysis suggests that inequality differences estimated from the HBS data, which are based solely on point estimates, should be made very cautiously. Even when comparing the estimates calculated for subperiods covering four or more years, there are many cases, especially for analysis of the subpopulations, when it is not known whether the observed and sometimes sizable changes in inequality result from random variation in the survey samples or from real movements in incomes of the population. Examples from our study include 18.8% change in QSR(0.8) index estimated for income distribution in Krakow during 1998–2008 or 10% change in the Gini index estimated for consumption distribution in Warsaw during 1998–2003, which are both statistically insignificant at the 0.01 level. This conclusion is all the more relevant for frequently made year-to-year inequality comparisons, which usually involve much smaller differences in the estimates.

References

Amiel Y., Cowell F., Polovin A. (1996), Inequality among the Kibbutzim, *Economica*, 63 (250), 63–85.

Bhattacharya D. (2007), Inference on Inequality from Household Survey Data, *Journal of Econometrics*, 137 (2), 674–707.

Biewen M. (2002), Bootstrap Inference for Inequality Mobility and Poverty Measurement, *Journal of Econometrics*, 108 (2), 317–342.

Biewen M., 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 (3), 371–383.

Binder D.A., Kovačević M.S. (1995), Estimating Some Measures of Income Inequality from Survey Data: an Application of the Estimating Equations Approach, *Survey Methodology*, 21 (2), 137–145.

Brandolini A. (2007), Measurement of Income Distribution in Supranational Entities: The Case of the European Union, in: S.P. Jenkins, J. Micklewright (eds.), *Inequality and Poverty Re-Examined*, Oxford University Press, New York.

Brandolini A., Smeeding T.M. (2008), Inequality: International Evidence, in: S.N. Durlauf, E.L. Blume, *The New Palgrave Dictionary of Economics*, 4, Basingstoke, Palgrave Macmillan.

Central Statistical Office (2008), *Household Budget Surveys in 2007*, CSO, Warszawa.

Cowell F.A. (1989), Sampling Variance and Decomposable Inequality Measures, *Journal of Econometrics*, 42(1), 27–42.

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

Cowell F.A. (2008), *Measuring Inequality*, forthcoming in Oxford University Press, http://darp.lse.ac.uk/papersDB/Cowell_measuringinequality3.pdf.

Cowell F.A., Flachaire E. (2007), Income Distribution and Inequality Measurement: The Problem of Extreme Values, *Journal of Econometrics*, 141 (2), 1044–1072.

Cowell F.A., Victoria-Feser M.-P. (1996), Robustness Properties of Inequality Measures, *Econometrica*, 64 (1), 77–101.

Cowell F.A., Victoria-Feser M.-P. (2007), Robust Lorenz Curves: a Semi-parametric Approach, *Journal of Economic Inequality*, 5 (1), 21–35.

Daras T., Zienkowski L., Żółkiewski Z. (2006), Zróżnicowanie dochodów i sfera ubóstwa w Polsce w latach 1993–2004, *Bank i Kredyt*, 37 (11–12), 27–45.

Davidson R. (2009), Reliable Inference for the Gini Index, *Journal of Econometrics*, 150 (1), 30–40.

Deaton A. (1997), *The Analysis of Household Surveys: a Microeconometric Approach to Development Policy*, Johns Hopkins University Press, Baltimore, MD.

Deville J.C. (1999), Variance Estimation for Complex Statistics and Estimators: Linearization and Residual Techniques, *Survey Methodology*, 25 (2), 193–203.

Dupuis D., Victoria-Feser M.-P. (2006), A Robust Prediction Error Criterion for Pareto Modelling of Upper Tails, *The Canadian Journal of Statistics*, 34 (4), 639–658.

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, Boston College Department of Economics.

Jenkins S.P., Biewen M. (2005), *SVYGEI SVYATK: Stata Module to Derive the Sampling Variances of Generalized Entropy and Atkinson Inequality Indices when Estimated from Complex Survey Data*, Statistical Software Components Archive S453601, Boston College Department of Economics.

Jenkins S.P., Van Kerm P. (2009), The Measurement of Economic Inequality, in: W. Salverda, B. Nolan, T.M. Smeeding (eds.), *Oxford Handbook on Economic Inequality*, Oxford University Press.

Keane M.P., Prasad E.S. (2002), Inequality, Transfers and Growth: New Evidence from the Economic Transition in Poland, *Review of Economics and Statistics*, 84 (2), 324–341.

Keane M.P., Prasad E.S. (2006), Changes in the Structure of Earnings during the Polish Transition, *Journal of Development Economics*, 80 (2), 389–427.

Kish L., Frankel M.R. (1974), Inference from Complex Samples, *Journal of the Royal Statistical Society B*, 36 (1), 1–37.

Kordos J., Lednicki B., Zyra M. (2002), The Household Sample Surveys in Poland, *Statistics in Transition*, 5 (4), 555–589.

Kovačević M.S., Binder D.A. (1997), Variance Estimation for Measures of Income Inequality and Polarization, *Journal of Official Statistics*, 13 (1), 41–58.

Langel M., Tillé Y. (2009), *Variance Estimation of the Quintile Share Ratio*, Communication au Journées Suisses de la Statistique, Genève, Août.

Levy H., Morawski L. (2007), *EUROMOD Country Report – Poland*, unpublished.

Maasoumi E. (1997), Empirical Analyses of Inequality and Welfare, in: M.H. Pesaran, P. Schmidt (eds.), *Handbook of Applied Econometrics: Microeconomics,* Blackwell, Oxford.

Milanović B. (1999), Explaining the Increase in Inequality during Transition, *Economics of Transition*, 7 (2), 299–341.

Mills J.A., Zandvakili S. (1997), Statistical Inference via Bootstrapping for Measures of Inequality, *Journal of Applied Econometrics*, 12 (2), 133–150.

Mitra P., Yemtsov R. (2006), *Increasing Inequality in Transition Economies: Is There More to Come?*, World Bank Policy Research Working Paper, 4007.

Newell A., Socha M. (2007), The Polish Wage Inequality Explosion, *Economics of Transition*, 15 (4), 733–758.

Osier G. (2009), Variance Estimation for Complex Indicators of Poverty and Inequality using Linearization Techniques, *Survey Research Methods*, 3 (3), 167–195.

StataCorp. (2009), *Stata Survey Data Reference Manual: Release 11*, Stata Corporation, College Station TX.

Szulc A. (2000), Economic Transition, Poverty, and Inequality: Poland in the 1990s, *Statistics in Transition*, 4 (6), 997–1017.

Szulc A. (2003), *Inequality During the Economic Transition in Poland: 1993–1999 Evidence*, unpublished manuscript.

Szulc A. (2006), *Poverty in Poland during the 1990s*: Are the Results Robust?, Review of Income and Wealth, 52 (3), 423–448.

Szulc A. (2008), Checking the Consistency of Poverty in Poland: 1997–2003 Evidence, *Post-Communist Economies*, 20 (1), 33–55.

UNICEF (2009), *TransMONEE 2009 Database*, UNICEF Regional Office for CEE/CIS, Geneva.

Van Kerm P. (2007), *Extreme Incomes and the Estimation of Poverty and Inequality Indicators from EU-SILC*, IRISS Working Paper Series, 2007-01, IRISS at CEPS/INSTEAD.

Acknowledgements

Most of the work for this paper was completed shortly before the untimely passing of my dear friend Krzysztof Kostro in August 2009.

We would like to thank two anonymous referees for very detailed comments and constructive suggestions that helped greatly to improve this paper. We also thank Bronislaw Lednicki (Central Statistical Office) for providing data on sample design for Household Budget Survey study. Any remaining errors are our own responsibility.

Appendix

Table 1

Distributions of disposable incomes and consumption expenditures in all Poland (the mean and various percentiles)

Year	P1	P10	P50	P90	P99	Mean
			Disposable incomes			
1998	249.4	525.1	980.7	1 803.2	3 670.6	1 122.3
1999	236.7	512.3	979.9	1 820.1	3 541.5	1 117.7
2000	214.0	488.6	950.9	1 821.9	3 775.2	1 106.6
2001	222.8	492.0	962.8	1 851.4	3 765.4	1 117.5
2002	208.1	471.1	932.6	1 798.7	3 677.3	1 086.3
2003	206.1	466.7	929.2	1 820.4	3 788.0	1 086.6
2004	196.1	440.2	906.5	1 802.5	3 918.2	1 073.1
2005	205.3	457.1	913.3	1 822.3	4 001.7	1 085.5
2006	236.8	511.9	1 004.1	1 981.0	4 191.1	1 185.9
2007	242.2	567.4	1 084.7	2 139.7	4 804.8	1 296.2
2008	274.1	612.0	1 172.3	2 311.0	4 956.8	1 389.8
			Consumption expenditures			
1998	289.9	464.1	844.9	1 622.1	3 506.7	1 002.2
1999	278.4	451.8	826.6	1 613.4	3 493.2	979.4
2000	262.4	442.3	819.8	1 618.3	3 416.9	974.2
2001	262.0	441.7	822.9	1 615.0	3 284.4	970.2
2002	259.2	433.6	813.2	1 623.0	3 417.1	969.5
2003	264.2	435.7	823.5	1 666.9	3 668.9	990.3
2004	245.6	424.2	822.5	1 696.5	3 545.8	991.4
2005	243.3	424.3	801.2	1 629.3	3 492.7	964.4
2006	280.8	460.7	866.7	1 753.7	3 607.8	1 037.5
2007	291.4	486.4	918.3	1 861.5	4 072.8	1 105.0
2008	305.4	511.6	965.2	1 995.6	4 291.4	1 172.9

Notes: In PLN per month per equivalised person (December 2008 price levels, 1 USD = 2.97 PLN and 1 Euro = 4.02 PLN). P1 is the first percentile, P10 is the tenth percentile, P50 is the median, P90 is the ninety percentile and P99 is the ninety ninth percentile.

Table 2

Distributions of disposable incomes and consumption expenditures in rural Poland (the mean and various percentiles)

Year	P1	P10	P50	P90	P99	Mean
			Disposable incomes			
1998	162.3	442.9	836.0	1 505.7	3 237.4	956.0
1999	153.2	428.2	834.0	1 496.7	3 052.4	943.5
2000	145.8	405.5	814.0	1 468.5	3 052.2	924.6
2001	164.7	405.9	820.2	1 497.1	3 128.4	929.0
2002	152.7	400.2	791.7	1 457.6	3 056.4	908.8
2003	156.0	397.4	778.8	1 438.7	2 933.4	886.3
2004	147.1	369.2	747.8	1 406.8	3 076.6	861.8
2005	145.6	390.1	769.3	1 459.9	3 336.5	896.2
2006	156.9	439.6	851.0	1 622.0	3 614.4	994.6
2007	154.6	486.8	924.0	1 759.9	4 142.0	1 089.8
2008	170.6	506.9	991.9	1 910.8	4 232.3	1 173.0
			Consumption expenditures			
1998	270.0	409.1	696.9	1 300.7	2 628.3	811.5
1999	249.5	391.4	678.9	1 294.1	2 778.9	799.8
2000	248.7	389.3	686.9	1 280.2	2 612.4	798.2
2001	239.8	383.8	687.0	1 292.1	2 663.9	801.8
2002	241.7	378.8	678.9	1 294.0	2 741.5	798.8
2003	244.1	382.2	685.0	1 316.8	2 864.7	806.0
2004	223.8	367.8	676.7	1 321.7	2 689.9	795.3
2005	221.1	379.9	681.1	1 342.6	2 671.9	800.8
2006	249.8	415.2	751.3	1 455.0	2 897.0	876.6
2007	262.3	432.1	791.9	1 543.7	3 054.1	926.1
2008	266.0	447.2	835.1	1 637.2	3 493.9	989.4

Notes: see Table 1.

Table 3

Distributions of disposable incomes and consumption expenditures in urban Poland (the mean and various percentiles)

Year	P1	P10	P50	P90	P99	Mean
			Disposable incomes			
1998	335.2	606.5	1 079.7	1 943.3	3 869.0	1 228.8
1999	332.0	606.3	1 090.1	1 964.5	3 703.5	1 231.5
2000	306.9	570.2	1 055.9	2 017.0	4 168.6	1 228.1
2001	305.1	572.9	1 075.9	2 023.7	3 911.6	1 236.4
2002	294.1	547.3	1 040.7	1 975.2	4 045.3	1 199.8
2003	268.2	539.0	1 044.1	2 019.7	4 140.3	1 215.1
2004	254.3	520.6	1 023.7	2 010.1	4 298.6	1 202.1
2005	254.3	525.3	1 019.8	1 997.8	4 319.9	1 199.3
2006	303.1	583.8	1 112.4	2 148.8	4 547.7	1 302.5
2007	333.7	647.5	1 195.9	2 316.0	5 017.6	1 411.5
2008	387.4	703.7	1 300.6	2 510.8	5 175.7	1 524.8
			Consumption expenditures			
1998	316.8	540.7	948.7	1 789.2	4 128.8	1 123.2
1999	318.7	528.8	937.7	1 757.8	3 792.0	1 095.5
2000	292.8	509.9	924.3	1 782.1	3 858.4	1 090.7
2001	291.3	509.9	924.0	1 772.3	3 578.1	1 080.6
2002	283.1	499.9	912.3	1 777.1	3 773.4	1 080.1
2003	286.5	497.7	930.0	1 838.8	4 078.8	1 111.0
2004	269.9	486.3	933.3	1 883.8	4 034.0	1 113.4
2005	270.6	472.6	887.6	1 779.8	3 819.0	1 063.8
2006	304.1	505.1	956.7	1 909.5	3 965.3	1 135.9
2007	322.7	536.0	1 009.5	2 042.3	4 526.6	1 215.8
2008	342.9	567.1	1 060.4	2 179.0	4 788.7	1 288.0

Notes: see Table 1.

Table 4

Distributions of disposable incomes and consumption expenditures in Warsaw (the mean and various percentiles)

Year	P1	P10	P50	P90	P99	Mean
			Disposable incomes			
1998	521.1	842.2	1 443.2	2 814.2	5 992.8	1 706.6
1999	548.0	875.1	1 515.5	2 833.6	6 807.6	1 783.4
2000	451.4	848.2	1 588.2	3 086.4	6 465.7	1 866.1
2001	481.7	829.8	1 557.7	2 977.6	6 466.7	1 822.8
2002	437.0	807.3	1 544.7	3 065.0	7 408.3	1 819.5
2003	449.4	844.0	1 619.1	3 359.3	6 098.4	1 929.9
2004	346.2	759.2	1 557.4	3 508.7	8 526.4	1 967.1
2005	367.9	710.9	1 446.3	3 151.0	7 106.7	1 803.5
2006	431.4	813.7	1 602.1	3 681.0	8 788.4	2 034.4
2007	519.3	913.1	1 758.1	4 186.8	9 902.7	2 354.5
2008	459.2	991.2	1 952.8	4 390.6	8 588.3	2 409.8
			Consumption expenditures			
1998	468.1	731.0	1 246.5	2 378.9	5 344.3	1 477.9
1999	476.6	760.8	1 284.4	2 355.4	4 462.0	1 469.4
2000	437.6	762.6	1 344.1	2 511.0	4 960.9	1 561.2
2001	485.8	740.5	1 305.5	2 486.4	4 526.6	1 507.8
2002	396.7	746.9	1 327.7	2 731.2	6 152.4	1 619.1
2003	412.3	755.1	1 434.7	2 959.5	6 000.9	1 727.2
2004	389.4	699.8	1 440.7	3 003.0	6 396.0	1 722.0
2005	301.2	633.8	1 227.4	2 634.6	5 403.8	1 495.1
2006	366.0	688.4	1 352.7	2 930.7	6 376.1	1 648.3
2007	422.1	750.4	1 468.8	3 340.4	6 969.5	1 887.2
2008	443.6	811.0	1 577.3	3 415.3	8 609.9	2 001.0

Notes: see Table 1.

Table 5

Distributions of disposable incomes and consumption expenditures in Krakow (the mean and various percentiles)

Year	P1	P10	P50	P90	P99	Mean
			Disposable incomes			
1998	501.1	774.2	1 255.0	2 334.2	4 320.9	1 432.6
1999	447.1	725.6	1 274.7	2 442.0	4 536.1	1 478.6
2000	355.0	657.5	1 190.8	2 111.7	4 100.9	1 344.3
2001	387.9	699.3	1 214.4	2 419.8	4 562.6	1 448.6
2002	233.2	609.0	1 174.4	2 223.8	5 175.3	1 389.3
2003	289.7	583.7	1 126.2	2 216.5	4 081.4	1 314.3
2004	325.2	646.2	1 236.2	2 428.2	5 224.3	1 453.7
2005	310.3	612.4	1 202.9	2 425.5	5 447.0	1 442.2
2006	302.7	713.1	1 259.3	2 278.6	4 220.5	1 415.5
2007	355.3	761.2	1 388.0	2 685.3	4 868.6	1 621.6
2008	483.6	832.9	1 530.0	3 021.3	5 163.0	1 775.6
			Consumption expenditures			
1998	516.4	685.1	1 090.4	1 985.5	3 520.7	1 253.0
1999	432.3	678.9	1 099.2	2 080.9	3 808.8	1 287.8
2000	432.6	633.6	1 073.7	1 880.4	3 321.5	1 208.0
2001	401.6	661.8	1 096.6	1 999.3	4 398.9	1 283.8
2002	273.7	566.8	1 101.2	2 044.6	3 894.3	1 256.1
2003	273.8	525.0	1 058.9	2 126.6	3 620.7	1 228.9
2004	316.2	609.1	1 178.7	2 356.0	5 821.9	1 379.5
2005	375.6	596.3	1 071.0	2 019.0	3 903.3	1 244.3
2006	393.3	606.5	1 119.1	2 288.7	4 092.4	1 285.2
2007	436.3	675.3	1 200.4	2 183.0	3 988.2	1 368.3
2008	437.9	722.2	1 278.8	2 513.3	5 231.6	1 522.3

Notes: see Table 1.

Table 6

Distributions of disposable incomes and consumption expenditures in Lodz (the mean and various percentiles)

Year	P1	P10	P50	P90	P99	Mean
			Disposable incomes			
1998	383.4	663.2	1 207.7	2 139.3	4 214.0	1 352.2
1999	350.0	666.4	1 210.8	2 118.2	3 945.4	1 338.9
2000	381.7	652.8	1 196.3	2 186.3	3 827.9	1 345.7
2001	226.4	556.7	1 126.0	2 192.2	3 695.1	1 283.8
2002	277.5	623.9	1 096.9	1 897.1	3 125.6	1 213.4
2003	303.0	609.8	1 104.1	2 214.2	4 160.9	1 283.3
2004	228.8	534.5	1 076.8	1 981.9	3 307.3	1 195.4
2005	232.2	539.3	1 055.8	1 961.6	4 176.1	1 230.5
2006	237.8	666.9	1 203.3	2 124.8	4 521.3	1 370.7
2007	207.5	652.8	1 231.6	2 170.7	4 430.6	1 359.7
2008	416.2	729.3	1 300.3	2 334.6	5 099.3	1 498.8
			Consumption expenditures			
1998	350.2	564.8	1 058.6	1 960.3	5 714.3	1 254.3
1999	311.8	597.9	1 022.8	1 914.1	4 428.8	1 203.0
2000	318.9	588.9	1 085.7	1 927.8	3 828.8	1 217.8
2001	257.0	551.4	995.6	1 832.9	2 980.1	1 122.0
2002	292.1	583.9	986.4	1 804.0	4 033.5	1 151.1
2003	303.4	581.4	1 019.2	1 987.2	4 020.2	1 183.5
2004	273.5	489.1	997.0	1 938.1	4 512.4	1 204.2
2005	297.2	520.4	984.4	2 007.0	3 818.0	1 162.8
2006	321.3	591.5	1 053.4	1 993.2	3 898.9	1 238.8
2007	298.1	571.6	1 062.9	2 118.9	5 625.3	1 277.2
2008	350.3	629.2	1 138.2	2 229.1	5 582.7	1340.8

Notes: see Table 1.