L. A. Karoly, C. Schröder

Fast methods for jackknifing inequality indices

The jackknife is a resampling method that uses subsets of the original database by leaving out one observation at a time from the sample. The paper develops fast algorithms for jackknifing inequality indices with only a few passes through the data. The number of passes is independent of the number of observations. Hence, the method provides an efficient way to obtain standard errors of the estimators even if sample size is large. We apply our method using micro data on individual incomes for Germany and the US.

Keywords: jackknife; resampling; sampling variability; inequality.

JEL classification: C81; C87; D3.

1. Introduction

When examining time-series changes in inequality or cross country differences in inequality, the measured changes are sometimes small. The delta method, bootstrapping and jackknifing are popular methods for performing statistical inference on inequality indices. The delta method is based on the asymptotic distribution of the index, whereas the jackknife and the bootstrap are distribution free resampling methods. The bootstrap samples with replacement from the original sample[1]. The jackknife uses subsets of the original database by leaving out one observation at a time from the sample[2].

The delta method, if applicable to an inequality index, has as advantages that it derives the asymptotic distribution using the central limit theorem, and that the computational burden to estimate the asymptotic variance consistently is low. However, the delta method is not free of limitations. For example, one difficulty arises if cross-sections of a panel survey are used as the empirical database, as it is the case for available micro datasets including the Luxembourg Income Study (LIS)[3], probably the most frequently used database for distributional analyses worldwide. Then for testing for inter-temporal changes in inequality indices the inter-temporal covariance structure of incomes must be considered. Further difficulties arise in the following contexts (see (Heshmati, 2004), and references therein for details): correlated data, correlations introduced in error terms[4], panel attrition, and inequality decomposition[5].

---

[1] For the validity of the bootstrap technique in inequality analyses see (Biewen, 2002).

[2] For the theoretical justification for the jackknife and other related resampling techniques see (Efron, 1982).

[3] http://www.lisdatacenter.org/.

[4] Modarres and Gastwirth (2006) show that regression estimates of the standard error for the Gini index can be biased «as it does not account for the correlation introduced in the error terms once the data are sorted» (p. 387).

[5] Decomposition analyses impose restrictions on the functional forms of regression models (see (Fields, Yoo, 2000), or (Morduch, Sicular, 2002)).

Aforementioned limitations of the delta method explain the interest in resampling methods[6]. The central advantage of the jackknife over other resampling methods such as the bootstrap is that it allows the replication of results. A disadvantage of standard jackknife procedures is that for large sample sizes the computational burden is substantial. This is because there are as many subsets as there are observations in the sample, and for each subset the jackknife statistic needs to be computed. This paper offers a solution. We provide fast algorithms, requiring only a few passes through the data, for jackknifing several popular inequality indices: coefficient of variation, variance of the logarithms, mean log deviation, Theil index, and Atkinson index[7]. Since the number of passes is independent of the number of observations, even for large samples the computational burden remains small.

To get an idea of the computational burden see Fig. 1. It charts the computer time in minutes for a standard procedure for jackknifing inequality indices as a function of sample size[8]. Computer time increases exponentially in sample size, and for a sample size of about 80 000 cases it already exceeds four hours. Since many comparative inequality analyses rely on data from several points in time, countries and income concepts, computing the jackknife for all results can easily take days or weeks. This is a serious limitation, especially for researchers who use data stored on external servers (e. g., the Luxembourg Income Study) and face limited processing power.

Section 2 explains our jackknife algorithm. Section 3 provides the results from an empirical application relying on data from the Luxembourg Income Study. Section 4 concludes. Derivations of all the formulas are provided in an Appendix.

2. Efficient jackknife procedures for inequality indices

The jackknife offers a conceptually simple way to estimate the precision of a statistic, see the pioneering (Tukey, 1958; Efron, 1982; Efron, Gong, 1983; Wolter, 1985). In the context of inequality measurement, we have a random sample of N observations on income, $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)$ and sampling weights, $\omega_1, \omega_2, \ldots, \omega_N$. Let $\theta = \theta(\boldsymbol{y})$ denote our measure of inequality. Let $\theta_{(i)} = \theta(y_1, y_2, \ldots, y_{i-1}, y_{i+1}, \ldots, y_N)$ denote the jackknife estimate of the same measure of inequality for the subset where the i-th observation has been deleted.

Following Wolter (1985), the jackknife estimate of the standard error of θ is

$$SE_\theta = \left(\frac{N-1}{N} \sum_{i=1}^{N} \frac{\omega_i}{\bar{\omega}} \left[\theta_{(i)} - \theta \right]^2 \right)^{0.5}, \qquad (1)$$

[6] For example, Biewen (2002) shows the validity of the bootstrap in a number of aforementioned contexts. Modarres and Gastwirth (2006) recommend bootstrap and jackknife as reliable methods for random samples as opposed to other methods that assume independence between errors when they are dependent (as it is the case for the Gini coefficient). Bhattacharya (2007) suggests techniques of asymptotic inference of the Gini coefficient based on process theory and the delta method. The resulting variance formula, however, is rather difficult to implement (Davidson, 2009).

[7] Algorithms for the Gini coefficient are provided in (Karagiannis, Kovacevic, 2000) and (Yitzhaki, 1991). Karoly (1989) derives jackknife procedures for calculating the between- and within-group inequality components of the variance of the logarithms, the mean log deviation, and the Theil index. Ogwang (2000) shows that it is also possible to obtain standard errors for the Gini index from OLS regression. Giles (2004) extends the regression-based approach to test hypotheses regarding the sensitivity of the Gini coefficient to changes in the data using seemingly unrelated regressions.

[8] We have used the STATA software package *inequal7.ado* on the following hardware: 64-bit system; 8 GB RAM; Core 2 Duo CPU, 3GHz. STATA code is available in the authors' working paper (Karoly, Schröder, 2014).

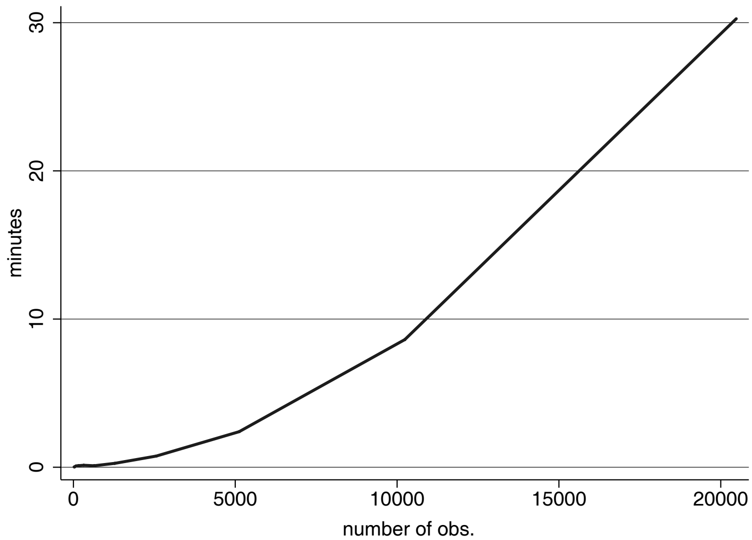

Fig. 1. Computer time and sample size

with $\bar{\omega} = \frac{1}{N}\sum_{i=1}^{N}\omega_i$ [9]. Computing the jackknife standard error estimate relies on the values of $\theta_{(i)}$, one jackknife statistic per subset. For large samples the computational burden to derive equation (1) seems to be large. However, as we will outline below, for standard inequality indices deriving the *N* values of $\theta_{(i)}$ requires just a few passes through the data. Hereby, the number of passes is independent of the number of sample observations, *N*.

The procedure is detailed below by means of the Theil index, and the variance of logarithms. The general idea of the procedure is to write the jackknife estimates as a function of statistics from the overall sample (i. e., as a function of θ, *N*, arithmetic or geometric mean) and a subset-specific correction factor that can be derived with a single run through the data. The procedure can be adapted to other inequality indices including indices of the generalized entropy class, and indices based on the variance or social-welfare functions (e. g. the Atkinson index).

We will make use of the following notation and definitions.

1. w_i denotes the normalized weight, $w_i = \omega_i \Big/ \frac{1}{N}\sum_{i=1}^{N}\omega_i$. Accordingly, $\sum_{i=1}^{N} w_i = N$.

2. $\bar{y}$ denotes the arithmetic mean of income, $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} w_i y_i$.

3. y^* denotes the geometric mean of income, $y^* = \exp\left(\frac{1}{N}\sum_{i=1}^{N} w_i \ln(y_i)\right)$. The natural logarithm of the geometric mean is denoted $\bar{x} = \ln(y^*) = \frac{1}{N}\sum_{i=1}^{N} w_i x_i$ with $x_i = \ln(y_i)$.

[9] An alternative method is to compute the squared differences between the jackknife statistics and their mean (see, for example, (Yitzhaki, 1991)).

2.1. Efficient jackknife procedure for the Theil index

The Theil index from the sample is

$$\theta_T = \frac{1}{N\bar{y}}\left(\sum_{i=1}^{N} w_i y_i \ln\left(y_i\right)\right) - \ln\left(\bar{y}\right). \quad (2)$$

The Theil index for the subset where the *i*-th observation has been deleted is

$$\theta_{T(i)} = \frac{1}{\left(N - w_i\right)\bar{y}_{(i)}}\left(\sum_{j\neq i} w_j y_j \ln\left(y_j\right)\right) - \ln\left(\bar{y}_{(i)}\right), \quad (3)$$

with $\bar{y}_{(i)}$ denoting the arithmetic mean of income from the subset

$$\bar{y}_{(i)} = \frac{N\bar{y} - w_i y_i}{N - w_i}. \quad (4)$$

The first step is to write $\theta_{T(i)}$ in terms of θ_T. Initially, from (3):

$$\theta_{T(i)} = \frac{1}{\left(N - w_i\right)\bar{y}_{(i)}}\left(\sum_{i=1}^{N} w_i y_i \ln\left(y_i\right)\right) - \frac{w_i}{\left(N - w_i\right)}\frac{y_i}{\bar{y}_i}\ln\left(y_i\right) - \ln\left(\bar{y}_{(i)}\right). \quad (5)$$

Rewriting equation (2) gives

$$\sum_{i=1}^{N} w_i y_i \ln\left(y_i\right) = \left[\theta_T + \ln\left(\bar{y}\right)\right] N\bar{y}, \quad (6)$$

and substituting (6) and (4) into (5) gives

$$\theta_{T(i)} = \frac{N\bar{y}}{N\bar{y} - w_i y_i}\left(\theta_T + \ln\left(\bar{y}\right)\right) - \frac{w_i y_i \ln\left(y_i\right)}{N\bar{y} - w_i y_i} - \ln\left(\frac{N\bar{y} - w_i y_i}{N - w_i}\right). \quad (7)$$

Equation (7) reveals that $\theta_{T(i)}$ can be expressed as a function of three statistics from the full sample, $N, \bar{y}$, and θ_T, and characteristics of the observation that is left out, w_i and y_i. Thus, after having calculated $N, \bar{y}$, and θ_T for the full sample, to compute all the jackknife statistics $\theta_{T(1)}, \ldots, \theta_{T(N)}$ takes a single pass through the data.

2.2. Efficient jackknife procedure for the variance of logarithms

Applying Bessel's correction[10], the variance of the logarithms from the sample is

$$\theta_{VL} = \frac{1}{N-1}\sum_{i=1}^{N} w_i \ln\left(\frac{y_i}{y^*}\right)^2 = \frac{1}{N-1}\sum_{i=1}^{N} w_i \left(x_i - \bar{x}\right)^2. \quad (8)$$

[10] Bessel's correction, the division in the variance formula by $N-1$ instead of by N, secures unbiasedness.

The variance of the logarithms for the subset where the i-th observation has been deleted is

$$\theta_{VL(i)} = \frac{1}{N-2}\sum_{j\neq i} w_{j(i)}\left(x_j - \overline{x}_{(i)}\right)^2, \tag{9}$$

with $\overline{x}_{(i)} = \frac{1}{N-w_i}\left[N\overline{X} - x_i w_i\right]$, and with $w_{j(i)} = \frac{w_j}{\left(N-w_i\right)/\left(N-1\right)}$ denoting re-weighted normalized weights. By means of the re-weighting the average of $w_{j(i)}$ over the subset where the i-th observation has been deleted equals unity. So, the analogue of the term $\frac{1}{N-1}$ in (8) in (9) is $\frac{1}{N-2}$. Substituting the definition of $w_{j(i)}$ in (9) gives

$$\theta_{VL(i)} = \frac{(N-1)}{(N-2)\left(N-w_i\right)}\sum_{j\neq i} w_j\left(x_j - \overline{x}_{(i)}\right)^2. \tag{10}$$

Initially, from (8):

$$\theta_{VL} = \frac{1}{N-1}\sum_{j\neq i} w_j\left(x_j - \overline{x}\right)^2 + \frac{1}{N-1} w_i\left(x_i - \overline{x}\right)^2. \tag{11}$$

Substituting $\overline{x} = \frac{1}{N}\left[\left(N-w_i\right)\overline{x}_{(i)} + x_i w_i\right]$ in (11) gives

$$\begin{aligned}\theta_{VL} &= \frac{1}{N-1}\sum_{j\neq i} w_j\left(x_j - \frac{1}{N}\left[\left(N-w_i\right)\overline{x}_{(i)} + x_i w_i\right]\right)^2 + \frac{1}{N-1} w_i\left(x_i - \overline{x}\right)^2 = \\ &= \frac{1}{N-1}\sum_{j\neq i} w_j\left(\underbrace{x_j - \frac{N}{N}\overline{x}_{(i)}}_{A} + \underbrace{\frac{w_i}{N}\overline{x}_{(i)} - \frac{w_i}{N}x_i}_{B}\right)^2 + \frac{1}{N-1} w_i\left(x_i - \overline{x}\right)^2.\end{aligned} \tag{12}$$

Equation (12) can be rewritten as

$$\begin{aligned}\theta_{VL} &= \underbrace{\frac{1}{N-1}\sum_{j\neq i} w_j\left(x_j - \overline{x}_{(i)}\right)^2}_{C} + \underbrace{\frac{2}{(N-1)}\sum_{j\neq i} w_j\left(\left(x_j - \overline{x}_{(i)}\right)\left(\frac{w_i}{N}\right)\left(\overline{x}_{(i)} - x_i\right)\right)}_{D} + \\ &+ \underbrace{\frac{1}{N-1}\sum_{j\neq i} w_j\left(\frac{w_i}{N}\overline{x}_{(i)} - \frac{w_i}{N}x_i\right)^2}_{E} + \frac{1}{N-1} w_i\left(x_i - \overline{x}\right)^2.\end{aligned} \tag{13}$$

The C-term on the right hand side of (13) can be rewritten as $\theta_{VL(i)}\frac{(N-2)\left(N-w_i\right)}{(N-1)^2}$. The D-term is zero since

$$D = \frac{2}{N-1}\frac{w_i}{N}\sum_{j\neq i} w_j\left(x_j - \overline{x}_{(i)}\right)\left(\overline{x}_{(i)} - x_i\right) = \frac{2w_i}{(N-1)N}\left(\overline{x}_{(i)} - x_i\right)\underbrace{\sum_{j\neq i} w_j\left(x_j - \overline{x}_{(i)}\right)}_{=0} = 0. \tag{14}$$

The E-term after some algebra becomes

$$E = \frac{1}{N-1}\frac{w_i^2}{\left(N-w_i\right)^2}\sum_{j\neq i} w_j\left(\bar{x}-x_i\right)^2 = \frac{1}{N-1}\frac{w_i^2}{\left(N-w_i\right)^2}\left(N-w_i\right)\left(\bar{x}-x_i\right)^2 = \\ = \frac{1}{N-1}\,\frac{w_i^2}{N-w_i}\left(\bar{x}-x_i\right)^2. \tag{15}$$

Substituting (14)–(15) in (13), the variance of the logarithms for the sample becomes

$$\theta_{VL} = \theta_{VL(i)}\frac{(N-2)\left(N-w_i\right)}{(N-1)^2} + \frac{1}{N-1}\frac{w_i^2}{N-w_i}\left(\bar{x}-x_i\right)^2 + \frac{w_i}{N-1}\left(x_i-\bar{x}\right)^2. \tag{16}$$

After some algebra, (16) becomes

$$\theta_{VL} = \theta_{VL(i)}\frac{(N-2)\left(N-w_i\right)}{(N-1)^2} + \frac{Nw_i}{(N-1)\left(N-w_i\right)}\left(\bar{x}-x_i\right)^2. \tag{17}$$

Solving (17) with respect to $\theta_{VL(i)}$ gives the desired expression for the jackknife estimator of the variance of the logarithms:

$$\theta_{VL(i)} = \theta_{VL}\frac{(N-1)^2}{(N-2)\left(N-w_i\right)} - \frac{Nw_i\left(N-1\right)}{\left(N-w_i\right)^2\left(N-2\right)}\left(\bar{x}-x_i\right)^2. \tag{18}$$

Equation (18) is the analogue of the jackknife estimator of the Theil index in equation (7): $\theta_{VL(i)}$ can be expressed as a function of statistics from the full sample ($N, \bar{x}$, and θ_{VL}) and the characteristics of the observation that is left out, w_i and x_i. Thus, after having calculated $N, \bar{x}$, and θ_{VL} for the full sample, computing $\theta_{VL(1)},\ldots,\theta_{VL(N)}$ takes a single pass through the data.

2.3. Efficient jackknife procedure for other inequality indices

Similar derivations as those explained in Sections 2.1 and 2.2 can be made for other inequality indices. Formulas for an efficient computation of the Atkinson index, θ_{A_ε} (with inequality aversion parameter $\varepsilon = 1$ and $\varepsilon = 2$), the mean log deviation, θ_{MLD}, and the coefficient of variation, θ_{CV}, are as follow:

$$\theta_{A_1(i)} = 1 - \frac{\exp\left[\frac{N}{N-w_i}\ln\left(y^*\right) - \frac{\ln\left(y_i\right)w_i}{N-w_i}\right]}{(N\bar{y}-w_i y_i)/(N-w_i)}, \tag{19}$$

$$\theta_{A_2(i)} = 1-(N-w_i)\Bigg/\left(\frac{N\bar{y}-w_i y_i}{\bar{y}\left(N-w_i\right)}\frac{N}{1-\theta_{A_2}} - \frac{w_i\left(N\bar{y}-w_i y_i\right)}{y_i\left(N-w_i\right)}\right), \tag{20}$$

$$\theta_{MLD(i)} = \frac{N}{N-w_i}\left[\theta_{MLD} - \ln\left(\bar{y}\right)\right] + \frac{w_i \ln\left(y_i\right)}{N-w_i} + \ln\left(\frac{N\bar{y} - y_i w_i}{N-w_i}\right), \quad (21)$$

$$\theta_{CV(i)} = \frac{\left(\theta_V \frac{(N-1)^2}{(N-2)(N-w_i)} - \frac{N w_i (N-1)}{(N-w_i)^2 (N-2)}\left(\bar{y} - y_i\right)^2\right)^{0.5}}{(N-w_i)/(N\bar{y} - y_i w_i)}. \quad (22)$$

Derivations of the formulas can be found in the Appendix. Again, after having calculated some basic statistics from the full sample, computing all the jackknife indices takes only a single pass through the data.

3. Empirical application

We have calculated the above inequality indices and their associated jackknife confidence intervals for distributions of disposable household incomes in the US and in Germany from the Luxembourg Income Study (LIS) database. For 40 countries and several years, the LIS provides representative micro-level information on private households' incomes and their demographics.

Our computations rely on the LIS household-level datasets. Household disposable income is our income concept. Household disposable income covers labor earnings, property income, and government transfers in cash minus income and payroll taxes. The Luxembourg Income Study spends enormous efforts in providing harmonized (standardized) income data, both in terms of conceptual content (variable definitions are comparable across datasets) and in terms of coding structure. Further information on the LIS data is provided online[11].

To adjust household incomes for differences in needs, we have deflated household disposable income by means of the square root equivalence scale. The square root equivalence scale is the number of household members to the power of 0.5. This gives the needs-adjusted equivalent income of the household. Household units are weighted by the frequency weights[12] times the number of household members. Our weighting procedure accommodates the principle of normative individualism that considers any person as important as any other. The so derived distribution depicts differences in living standards, captured by differences in equivalent incomes, among individuals (Bönke, Schröder, 2012).

We have removed household observations with missing information or with negative values of disposable income. Moreover, to avoid outlier-driven biases of inequality estimates, we use trimmed data with the one percent observations with the highest and with the lowest incomes being discarded.

It has taken a few seconds to obtain all the results presented in Table 1. The Table is split in two panels. The upper panel provides the results for the US, the lower panel provides the results

[11] For details see http://www.lisdatacenter.org/our-data/lis-database/documentation/.

[12] Household frequency weights are provided from the national data providers. For example, the German frequency weights come from the German data provider, the Socio-economic Panel. These weights make the sample representative of the total (covered) national population (usually, collective or institutionalized households are left out).

for Germany[13]. In the US, the results cover the period 1991–2010; in Germany, the results cover the period 1994–2010. For every country-period combination, the Table provides the point estimates of the inequality indices along with their upper and lower bounds of 95 percent confidence intervals, CI_{θ}^{lo} and CI_{θ}^{hi}, derived from the jackknife statistics.

Table 1. Inequality indices

Country	Year	Atkinson ($\varepsilon = 1$)			Atkinson ($\varepsilon = 2$)			Mean log deviation		
		$CI_{\theta_{A1}}^{lo}$	θ_{A1}	$CI_{\theta_{A1}}^{hi}$	$CI_{\theta_{A2}}^{lo}$	θ_{A2}	$CI_{\theta_{A2}}^{hi}$	$CI_{\theta_{MLD}}^{lo}$	θ_{MLD}	$CI_{\theta_{MLD}}^{hi}$
US	1991	0.162	0.166	0.169	0.329	0.337	0.345	0.177	0.181	0.186
	1997	0.177	0.181	0.185	0.348	0.357	0.366	0.195	0.199	0.204
	2000	0.173	0.177	0.180	0.340	0.348	0.356	0.190	0.194	0.199
	2004	0.179	0.183	0.186	0.361	0.371	0.380	0.197	0.202	0.206
	2007	0.185	0.188	0.191	0.363	0.370	0.377	0.204	0.208	0.212
	2010	0.193	0.197	0.201	0.402	0.411	0.421	0.215	0.219	0.224
DE	1994	0.088	0.095	0.102	0.175	0.188	0.200	0.093	0.100	0.107
	2000	0.088	0.093	0.098	0.174	0.185	0.196	0.092	0.098	0.103
	2004	0.098	0.106	0.114	0.184	0.203	0.222	0.103	0.112	0.121
	2007	0.102	0.111	0.120	0.193	0.210	0.226	0.107	0.117	0.127
	2010	0.103	0.110	0.117	0.198	0.212	0.225	0.109	0.116	0.124
		Theil index			Variance of logs			Coefficient of variation		
		$CI_{\theta_T}^{lo}$	θ_T	$CI_{\theta_T}^{hi}$	$CI_{\theta_{VL}}^{lo}$	θ_{VL}	$CI_{\theta_{VL}}^{hi}$	$CI_{\theta_{CV}}^{lo}$	θ_{CV}	$CI_{\theta_{CV}}^{hi}$
US	1991	0.158	0.161	0.165	0.396	0.408	0.419	0.574	0.581	0.587
	1997	0.180	0.184	0.189	0.422	0.435	0.447	0.637	0.646	0.654
	2000	0.177	0.181	0.185	0.410	0.421	0.432	0.633	0.643	0.653
	2004	0.178	0.182	0.185	0.439	0.452	0.464	0.625	0.633	0.640
	2007	0.188	0.192	0.196	0.445	0.456	0.466	0.653	0.661	0.669
	2010	0.189	0.192	0.196	0.494	0.508	0.522	0.639	0.646	0.652
DE	1994	0.090	0.097	0.104	0.191	0.207	0.222	0.437	0.456	0.475
	2000	0.090	0.095	0.099	0.190	0.203	0.216	0.439	0.451	0.463
	2004	0.103	0.111	0.119	0.204	0.226	0.248	0.480	0.500	0.519
	2007	0.108	0.118	0.129	0.213	0.234	0.254	0.493	0.522	0.551
	2010	0.107	0.114	0.122	0.221	0.238	0.254	0.482	0.504	0.525

Note. Data from Luxembourg Income Study.

We comment on the US first. An examination of the statistics shows a significant rise of inequality over the observation period: the point estimate of the Theil index increases from 0.161 in 1991 to 0.192 in 2010, and the confidence intervals are clearly distinct: $[0.158; 0.165]$ vs.

[13] The LIS data for Germany are based on the German Socio-Economic Panel Study (SOEP).

$[0.189; 0.196]$. However, some inter-temporal changes in inequality for this sample are not statistically significant (e. g. 1997–2000; 2000–2004; 2004–2007).

For Germany, we also see a significant rise of inequality over the observation period. This is due to a prominent rise of inequality between 2000 and 2004. The inter-temporal comparisons before the rise (1994–2000) and after the rise (2004–2007 and 2007–2010) indicate no significant changes in inequality.

Comparing inequality levels in the US and Germany there is significantly more inequality in the US. The result holds for all six inequality indices and all the observed points in time[14].

Finally, as empirical illustration, we compare the jackknife confidence intervals for the Theil index with the normal 95% confidence intervals from the bootstrap and from asymptotic variances[15]. The results are summarized in Table 2.

Table 2. Alternative confidence intervals for the Theil index

Country	Year	Jackknife		Bootstrap		Asymptotic	
		$CI^{lo}_{\theta_T}$	$CI^{hi}_{\theta_T}$	$CI^{lo}_{\theta_T}$	$CI^{hi}_{\theta_T}$	$CI^{lo}_{\theta_T}$	$CI^{hi}_{\theta_T}$
US	1991	0.158	0.165	0.159	0.163	0.159	0.163
	1997	0.180	0.189	0.182	0.187	0.182	0.187
	2000	0.177	0.185	0.178	0.184	0.178	0.184
	2004	0.178	0.185	0.179	0.184	0.179	0.184
	2007	0.188	0.196	0.189	0.195	0.189	0.195
	2010	0.189	0.196	0.190	0.195	0.190	0.195
DE	1994	0.090	0.104	0.093	0.102	0.093	0.101
	2000	0.090	0.099	0.091	0.098	0.091	0.098
	2004	0.103	0.119	0.107	0.116	0.107	0.115
	2007	0.108	0.129	0.113	0.124	0.113	0.124
	2010	0.107	0.122	0.110	0.119	0.110	0.119

Note. Data from Luxembourg Income Study.

The general message from Table 2 is that the confidence intervals from the three methods are very close. Indeed, confidence intervals from the bootstrap and from asymptotic variances, for many years, coincide. Inter-temporal and cross-country comparisons are insensitive to the chosen methods. The inter-temporal comparisons indicate a significant rise in inequality for the US between 1991 and 1997 and again between 2004 and 2007 (Germany: between 2000 and 2004). The cross-country comparisons indicate that inequality is significantly higher in the US.

---

14 We have executed our empirical analysis using the alternative formulation of the standard error introduced in footnote 4. It did not change our conclusions since confidence intervals changed very tittle.

15 The bootstrap confidence intervals have been obtained with the STATA ado-package *ineqerr* with 1000 repetitions. The confidence intervals from asymptotic variances have been obtained with the STATA ado-package *geivars*, with the underlying formulas coming from (Cowell, 1989). Results for other inequality indices can be provided on request.

4. Conclusion

This paper has outlined a procedure to obtain jackknife estimates for several inequality indices with only a few passes through the data. The number of passes is independent of the number of observations: After having computed some statistics from the overall sample, computing all the jackknife indices takes only a single pass through the data. Hence, the method provides an efficient way to get standard errors of the estimators even if sample size is large.

We have applied our method using data from the Luxembourg Income Study to evaluate the statistical significance of inter-temporal inequality in Germany and the US, and also to evaluate cross country differences in inequality levels.

References

Bhattacharya D. (2007). Inference on inequality from household survey data. *Journal of Econometrics*, 137, 674–707.

Biewen M. (2002). Bootstrap inference for inequality, mobility and poverty measurement. *Journal of Econometrics*, 108, 317–342.

Bönke T., Schröder C. (2012). Country inequality rankings and conversion schemes. *Economics: The Open-Access, Open-Assessment E-Journal*, 6, 2012–2028.

Cowell F. A. (1989). Sampling variance and decomposable inequality measures. *Journal of Econometrics*, 42, 27–41.

Davidson R. (2009). Reliable inference for the Gini index. *Journal of Econometrics*, 150, 30–40.

Efron B. (1982). *The jackknife, the bootstrap and other resampling plans.* Society for Industrial and Applied Mathematics, Philadelphia PA.

Efron B., Gong G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37, 36–48.

Fields G. S., Yoo G. (2000). Falling labour income inequality in Korea's economic growth: Patterns and underlying causes. *Review of Income and Wealth*, 46, 139–159.

Giles D. (2004). Calculating a standard error for the Gini coefficient: Some further results. *Oxford Bulletin of Economics and Statistics*, 66, 425–433.

Heshmati A. (2004). A review of decomposition of income inequality. *IZA Discussion Paper*, 1221.

Karagiannis E., Kovacevic M. (2000). Practitioners corner — a method to calculate the jackknife variance estimator for the Gini coefficient. *Oxford Bulletin of Economics and Statistics*, 62, 119–122.

Karoly L. A. (1989). Computing standard errors for measures of inequality using the jackknife. *Unpublished manuscript*.

Karoly L. A., Schröder C. (2014). Fast methods for jackknifing inequality indices. *SOEPpapers on Multidisciplinary Panel Data Research* No. 643, DIW Berlin.

Modarres R., Gastwirth J. L. (2006). A cautionary note on estimating the standard error of the Gini index of inequality. *Oxford Bulletin of Economics and Statistics*, 68, 385–390.

Morduch J., Sicular T. (2002). Rethinking inequality decomposition, with evidence from rural China. *The Economic Journal*, 112, 93–106.

Ogwang T. (2000). A convenient method of computing the Gini index and its standard error. *Oxford Bulletin of Economics and Statistics*, 62, 123–129.

Tukey J. W. (1958). Bias and confidence in not quite large samples. *Annals of Mathematical Statistics*, 29, 614–623.

Wolter K. (1985). *Introduction to variance estimation*. Springer, New York.

Yitzhaki S. (1991). Calculating jackknife variance estimators for parameters of the Gini method. *Journal of Business and Economic Statistics*, 9, 235–239.

Appendix

Derivation of jackknife formulas

Mean log deviation (Entropy 0)

$$\theta_{MLD} = \frac{1}{N}\sum_{i=1}^{N} w_i \ln\left(\frac{\bar{y}}{y_i}\right) = -\frac{1}{N}\sum_{i=1}^{N} w_i \ln(y_i) + \ln(\bar{y}). \qquad (1^{MLD})$$

$$\theta_{MLD(i)} = -\frac{1}{N-w_i}\sum_{j\neq i} w_j \ln(y_i) + \ln(\bar{y}). \qquad (2^{MLD})$$

From (2^{MLD}):

$$\theta_{MLD(i)} = -\frac{1}{N-w_i}\left[\sum_{j\neq i} w_j \ln(y_j) + w_i \ln(y_i)\right] + \frac{w_i \ln(y_i)}{N-w_i} + \ln(\bar{y}_{(i)}), \qquad (3^{MLD})$$

$$\theta_{MLD(i)} = -\frac{1}{N-w_i}\left[\sum_{i=1}^{N} w_i \ln(y_i)\right] + \frac{w_i \ln(y_i)}{N-w_i} + \ln(\bar{y}_{(i)}). \qquad (4^{MLD})$$

Substituting $-N\left[\theta_{MLD} - \ln(\bar{y})\right] = \sum_{i=1}^{N} w_i \ln(y_i)$ from (1^{MLD}) gives

$$\theta_{MLD(i)} = -\frac{1}{N-w_i}\left[-N\left[\theta_{MLD} - \ln(\bar{y})\right]\right] + \frac{w_i \ln(y_i)}{N-w_i} + \ln(\bar{y}_{(i)}). \qquad (5^{MLD})$$

Substituting $\bar{y}_{(i)}$ by $\dfrac{N\bar{y} - y_i w_i}{N-w_i}$ gives

$$\theta_{MLD(i)} = \frac{N}{N-w_i}\left[\theta_{MLD} - \ln(\bar{y})\right] + \frac{w_i \ln(y_i)}{N-w_i} + \ln\left(\frac{N\bar{y} - y_i w_i}{N-w_i}\right). \qquad (6^{MLD})$$

Atkinson Index

The general form of the Atkinson index is $\theta_{A_\varepsilon} = 1 - \left[\frac{1}{N}\sum_{i=1}^{N} w_i \left(\bar{y}/y_i\right)^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}$. Below we derive the jackknife formulas for two prominent case of the inequality aversion parameter, ε.

Inequality aversion parameter $\varepsilon = 1$

$$\theta_{A_1} = 1 - \frac{y^*}{\bar{y}} = 1 - \frac{\exp\left[\frac{1}{N}\sum_{i=1}^{N} w_i \ln(y_i)\right]}{\bar{y}}, \qquad (1^{A_1})$$

$$\theta_{A_1(i)} = 1 - \frac{\exp\left[\frac{1}{N - w_i}\sum_{j\neq i} w_j \ln\left(y_j\right)\right]}{\bar{y}_{(i)}}. \qquad \left(2^{A_1}\right)$$

Expansion of the term in brackets in the numerator with $\frac{\ln\left(y_i\right)w_i}{N - w_i} - \frac{\ln\left(y_i\right)w_i}{N - w_i}$, and substitution of $\bar{y}_{(i)}$ by $\frac{N\bar{y} - y_i w_i}{N - w_i}$ gives

$$\theta_{A_1(i)} = 1 - \frac{\exp\left[\frac{N}{N - w_i}\left(\frac{1}{N}\sum_{i=1}^{N} w_i \ln\left(y_i\right)\right) - \frac{\ln\left(y_i\right)w_i}{N - w_i}\right]}{\frac{N\bar{y} - w_i y_i}{N - w_i}}. \qquad \left(3^{A_1}\right)$$

Substitution of the term $\frac{1}{N}\sum_{i=1}^{N} w_i \ln\left(y_i\right)$ (log of the geometric mean of income from the full sample) by $\ln\left(y^*\right)$ gives

$$\theta_{A_1(i)} = 1 - \frac{\exp\left[\frac{N}{N - w_i}\ln\left(y^*\right) - \frac{\ln\left(y_i\right)w_i}{N - w_i}\right]}{\frac{N\bar{y} - w_i y_i}{N - w_i}}. \qquad \left(4^{A_1}\right)$$

Inequality aversion parameter $\varepsilon = 2$

$$\theta_{A_2} = 1 - \frac{N}{\sum_{i=1}^{N} w_i \frac{\bar{y}}{y_i}}, \qquad \left(1^{A_2}\right)$$

$$\theta_{A_2(i)} = 1 - \frac{N - w_i}{\sum_{j\neq i} w_j \frac{\bar{y}_{(i)}}{y_j}}. \qquad \left(2^{A_2}\right)$$

Expansion of the denominator with $w_i \frac{\bar{y}_{(i)}}{\bar{y}}\frac{\bar{y}}{y_i} - w_i \frac{\bar{y}_{(i)}}{\bar{y}}\frac{\bar{y}}{y_i}$ and rewriting the sum gives

$$\theta_{A_2(i)} = 1 - \frac{N - w_i}{\left(\frac{\bar{y}_{(i)}}{\bar{y}}\sum_{j\neq i} w_j \frac{\bar{y}}{y_j}\right) + w_i \frac{\bar{y}_{(i)}}{\bar{y}}\frac{\bar{y}}{y_i} - w_i \frac{\bar{y}_{(i)}}{\bar{y}}\frac{\bar{y}}{y_i}}, \qquad \left(3^{A_2}\right)$$

$$\theta_{A_2(i)} = 1 - \frac{N - w_i}{\left(\frac{\bar{y}_{(i)}}{\bar{y}}\sum_{i=1}^{N} w_i \frac{\bar{y}}{y_i}\right) - w_i \frac{\bar{y}_{(i)}}{\bar{y}}\frac{\bar{y}}{y_i}}. \qquad \left(4^{A_2}\right)$$

From $\theta_{A_2} = 1 - \frac{N}{\sum_{i=1}^{N} w_i \bar{y}/y_i}$ it follows that $\sum_{i=1}^{N} w_i \bar{y}/y_i = N \big/ \left(1-\theta_{A_2}\right)$, and replacement of the sum in the denominator gives

$$\theta_{A_2(i)} = 1 - \frac{N - w_i}{\frac{\bar{y}_{(i)}}{\bar{y}} \frac{N}{1-\theta_{A_2}} - \left(\frac{w_i \bar{y}_{(i)}}{y_i}\right)}. \qquad (5^{A_2})$$

Finally, substitution of $\bar{y}_{(i)}$ by $\frac{N\bar{y} - y_i w_i}{N - w_i}$ gives

$$\theta_{A_2(i)} = 1 - \frac{N - w_i}{\frac{N\bar{y} - w_i y_i}{\bar{y}\left(N - w_i\right)} \frac{N}{1-\theta_{A_2}} - \frac{w_i\left(N\bar{y} - w_i y_i\right)}{y_i\left(N - w_i\right)}}. \qquad (6^{A_2})$$

Variance and Coefficient of Variation

$$\theta_V = \frac{1}{N-1}\sum_{i=1}^{N} w_i\left(y_i - \bar{y}\right)^2, \qquad (1^V)$$

$$\theta_{V(i)} = \frac{(N-1)}{(N-2)\left(N-w_i\right)}\sum_{j\neq i} w_j\left(y_j - \bar{y}_{(i)}\right)^2. \qquad (2^V)$$

Rewriting of θ_V gives

$$\theta_V = \frac{1}{N-1}\sum_{j\neq i} w_j\left(y_j - \bar{y}_{(i)}\right)^2 + \frac{1}{N-1} w_i\left(y_i - \bar{y}\right)^2. \qquad (3^V)$$

Substituting $\bar{y} = \frac{1}{N}\left[\left(N - w_i\right)\bar{y}_{(i)} + y_i w_i\right]$ and reorganizing in analogy to the variance of the logarithms gives

$$\theta_V = C + D + E + \frac{1}{N-1} w_i\left(y_i - \bar{y}\right)^2, \qquad (4^V)$$

$$C = \frac{1}{N-1}\sum_{j\neq i} w_j\left(y_j - \bar{y}_{(i)}\right)^2 = \theta_{V(i)}\frac{(N-2)\left(N-w_i\right)}{(N-1)^2}, \qquad (5^V)$$

$$D = \frac{2}{N-1}\frac{w_i}{N}\sum_{j\neq i} w_j\left(y_j - \bar{y}_{(i)}\right)\left(\bar{y}_{(i)} - y_i\right) = \frac{2w_i}{(N-1)N}\left(\bar{y}_{(i)} - y_i\right)\underbrace{\sum_{j\neq i} w_j\left(y_j - \bar{y}_{(i)}\right)}_{=0} = 0, \qquad (6^V)$$

$$E = \frac{1}{N-1}\left(\frac{w_i}{N}\right)^2\sum_{j\neq i} w_j\left(\bar{y}_{(i)} - y_i\right)^2. \qquad (7^V)$$

Analogously to θ_{VL} we can rewrite $\left(7^V\right)$ as

$$E = \frac{1}{N-1}\frac{w_i^2}{N-w_i}\left(\bar{y}-y_i\right)^2. \qquad \left(8^V\right)$$

Substitution $\left(5^V\right),\left(6^V\right)$, and $\left(8^V\right)$ in $\left(4^V\right)$ gives

$$\theta_V = \theta_{V(i)}\frac{(N-2)\left(N-w_i\right)}{\left(N-1\right)^2} + \frac{1}{N-1}\frac{w_i^2}{N-w_i}\left(\bar{y}-y_i\right)^2 + \frac{w_i}{N-1}\left(y_i-\bar{y}\right)^2. \qquad \left(9^V\right)$$

Analogously to θ_{VL} we can rewrite $\left(9^V\right)$ as

$$\theta_V = \theta_{V(i)}\frac{(N-2)\left(N-w_i\right)}{\left(N-1\right)^2} + \frac{Nw_i}{(N-1)\left(N-w_i\right)}\left(\bar{y}-y_i\right)^2. \qquad \left(10^V\right)$$

Solving $\left(10^V\right)$ for $\theta_{V(i)}$ gives

$$\theta_{V(i)} = \theta_V\frac{\left(N-1\right)^2}{(N-2)\left(N-w_i\right)} - \frac{Nw_i\left(N-1\right)}{\left(N-w_i\right)^2\left(N-2\right)}\left(\bar{y}-y_i\right)^2. \qquad \left(11^V\right)$$

The coefficient of variation is defined as

$$\theta_{CV} = \left(\theta_V\right)^{0.5}\Big/\bar{y}. \qquad \left(1^{CV}\right)$$

Hence:

$$\theta_{CV(i)} = \left(\theta_{V(i)}\right)^{0.5}\Big/\bar{y}_{(i)}. \qquad \left(2^{CV}\right)$$

Substitution of $\theta_{V(i)} = \theta_V\frac{\left(N-1\right)^2}{(N-2)\left(N-w_i\right)} - \frac{Nw_i\left(N-1\right)}{\left(N-w_i\right)^2\left(N-2\right)}\left(\bar{y}-y_i\right)^2$

and of $\bar{y}_{(i)} = \frac{1}{\left(N-w_i\right)}\left(N\bar{y}-y_iw_i\right)$ in $\left(2^{CV}\right)$ gives

$$\theta_{CV(i)} = \frac{\left(\theta_V\frac{\left(N-1\right)^2}{(N-2)\left(N-w_i\right)} - \frac{Nw_i\left(N-1\right)}{\left(N-w_i\right)^2\left(N-2\right)}\left(\bar{y}-y_i\right)^2\right)^{0.5}}{\frac{1}{\left(N-w_i\right)}\left(N\bar{y}-y_iw_i\right)}. \qquad \left(3^{CV}\right)$$