The Economic Journal, **117** (*November*), F589–F603. © The Author(s). Journal compilation © Royal Economic Society 2007. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

HOW CLOSELY DO TOP INCOME SHARES TRACK OTHER MEASURES OF INEQUALITY?*

Andrew Leigh

In recent years, researchers have used taxation statistics to estimate the share of total income held by the richest groups, such as the top 10% or the top 1%. Compiling a standardised top income shares dataset for 13 developed countries, I find that there is a strong and significant relationship between top income shares and broader inequality measures, such as the Gini coefficient. This suggests that panel data on top income shares may be a useful substitute for other measures of inequality over periods when alternative income distribution measures are of low quality, or unavailable.

Since Adam Smith, economists have devoted considerable attention to the causes and effects of inequality.[1] Attempting to explain changes in income distribution, economists have considered the impact of unionisation, trade, immigration, inflation, family structure, the age profile of the population, technological change, compulsory schooling, minimum wages and progressive taxation, to name but a few. Inequality has also found itself on the right-hand side of many regressions. Researchers have investigated whether inequality affects growth, consumption, saving, infant mortality, height, residential segregation, happiness, trust, crime and political polarisation.[2]

However, much of the empirical research on income distribution has been plagued by a lack of high-quality data. Inequality measures are sometimes compared to one another despite the fact that they differ in their choice of reference group (individual, family or household), in the type of inequality being measured (income or expenditure), in the way that income is adjusted for family size and in whether the estimates take account of income taxation. Yet using more comparable estimates of income distribution, such as those from the Luxembourg Income Study (LIS), often means a substantial reduction in sample size.

This article considers an alternative source of data on inequality: measures of the income share held by the richest x% of the population, derived from tax return data. In recent years, estimates of top income shares for several developed countries have become available. Here, I consider the top incomes estimates available for 13 countries – Australia, Canada, France, Germany, Ireland, Japan, the Netherlands, New Zealand, Spain, Switzerland, Sweden, the UK and the US. What issues of comparability arise in using these data, as compared with other inequality data? How closely do they track the

* This article builds on the work of Facundo Alvaredo, Tony Atkinson, Fabien Dell, Chiaki Moriguchi, Brian Nolan, Thomas Piketty, Jesper Roine, Emmanuel Saez, Wiemer Salverda, Michael Veall, and Daniel Waldenström, who have painstakingly used taxation statistics and other historical data to estimate top income shares for the countries analysed herein. In addition, I am grateful to Tony Atkinson, Ian Irvine, Thomas Piketty, Emmanuel Saez and Daniel Waldenström for valuable comments on earlier drafts. Elena Varganova provided outstanding research assistance. The dataset of top income shares may be downloaded from http://econrsss.anu.edu.au/~aleigh/.

[1] Gilbert (1997) has discussed Adam Smith's writings on poverty and inequality in more detail.

[2] For a good summary of the scope of the field, see *Journal of Economic Inequality*, vol. 1, pp. 101–2 (2003).

income distribution as a whole? And how might they be used by researchers keen to learn more about the causes and effects of inequality?

The remainder of this article is structured as follows. Section 1 discusses the main data quality issues arising from the use of existing inequality datasets and top incomes data. Section 2 discusses and analyses the association between top income shares and other measures of inequality, and the final section concludes.

1. Data Quality

1.1 *Existing Inequality Datasets*

Over the past decade, most researchers studying inequality across countries have used one of three datasets: a database constructed by Deininger and Squire (1996), containing 2,632 Gini coefficients for 138 countries; the World Income Inequality Database (WIID), a more recent database compiled by the United Nations University and the World Institute for Development Economics Research, which contains 4,664 Gini coefficients for 154 countries; and the LIS, containing 143 measures of inequality for 30 countries.[3]

While the Deininger and Squire database and WIID have the advantage of extensive coverage across countries and over time, they also have the drawback that their measures of inequality are frequently not comparable with one another. In a seminal paper, Atkinson and Brandolini (2001) noted major problems arising from the use of these databases. They observed substantial comparability problems with the database, warned against the practice of researchers merely using the 'high quality' subset of the database and the use of dummy variable corrections to account for differences between measures of expenditure and income inequality.[4] Atkinson and Brandolini showed that certain inter-country and intra-country studies based upon the much-used Deininger and Squire (1996) database were not robust to measuring inequality using a different dataset that employed a consistent methodology for measuring inequality.[5]

For cross-country studies of inequality in developed countries, Atkinson and Brandolini advocate making greater use of the LIS, on the basis that it employs a consistent methodology across countries for measuring income and calculating inequality. Yet this smaller sample size comes at a cost – with 143 observations, the LIS is less than one-tenth the size of the WIID. Moreover, the LIS has very limited coverage prior to 1980.[6] These factors limit the scope for careful econometric studies, particularly if one wishes to include a country-specific dummy in the regression, or investigate the causes and effects of inequality over the very long run.

[3] Deininger and Squire dataset downloaded from http://www.worldbank.org on 20 December 2004. WIID is version 2*a* (July 2005), downloaded from http://www.wider.unu.edu/wiid/wiid.htm. LIS data from http://www.lisproject.org/keyfigures/ineqtable.htm (file current as of 21 September 2006). In the LIS, there are two 1984 observations for France: I use the one from the Household Budget Survey (FR84B).

[4] Deininger and Squire identify a 'high quality' subset of their database, consisting of 693 observations from 116 countries, which they label 'accept'.

[5] Greater concern over the quality of inequality measures appears to have penetrated economics to some degree, but uncritical use of Deininger and Squire's dataset remains common in other disciplines. See for example Fearon and Laitin (2003).

[6] Although the earliest observation in the LIS is for 1969, only 13 observations appear prior to 1980, so the dataset essentially covers the 1980s and 1990s.

© The Author(s). Journal compilation © Royal Economic Society 2007

1.2. *Top Incomes*

Can top incomes data help to fill the void? Beginning with the work by Piketty (2001) on the long-run distribution of top incomes in France, top incomes series have now been developed for 13 developed countries. These are Australia (Atkinson and Leigh, 2007*a*), Canada (Saez and Veall, 2005), France (Piketty, 2001, 2003, 2007), Germany (Dell, 2005, 2007), Ireland (Nolan, 2007), Japan (Moriguchi and Saez, 2006), the Netherlands (Salverda and Atkinson 2007), New Zealand (Atkinson and Leigh, 2005), Spain (Alvaredo and Saez, 2006), Sweden (Roine and Waldenström, 2006), Switzerland (Dell, 2005; Dell *et al.*, 2007), the UK (Atkinson, 2005, 2007*b*) and the US (Piketty and Saez, 2001, 2003). Estimates are also available for the world's three largest developing nations: China (Piketty and Qian, 2006), India (Banerjee and Piketty, 2005) and Indonesia (Leigh and van der Eng, 2007). Others are presently preparing series for Argentina, Denmark, Finland, Italy and Norway.

Although earlier studies (including the seminal work of Kuznets, 1953, 1955) made use of taxation data to measure inequality, much of this prior literature suffered from the problem that its estimates were representative only of taxpayers, and not of the entire population. What distinguishes the recent literature is the use of external sources to produce the population and personal income control totals. Because the recent studies take into account the incomes of non-filers, their estimates of top income shares are more precise than those that preceded them.

Here, I focus on estimates of top incomes shares that have been prepared for developed nations. The main rationale for this exclusion is the greater reliability of taxation statistics in this group of countries than in China, India and Indonesia, where tax evasion is a more significant problem. Naturally, tax evasion may also affect estimates of top income shares in developed nations. For example, Alvaredo and Saez (2006) regard estimates of Spanish top incomes prior to 1981 as unreliable due to widespread tax evasion, so I only present data for Spain from 1981 onwards. This comparability exercise is designed to complement the work of Atkinson and Piketty (2007).[7] Data sources are set out in Appendix Table A1.

1.3. *Problems of Comparability*

Deriving income distribution measures from taxation data is not without its complications. The most severe of these is that individuals have a strong incentive to underreport income to the tax authorities. If the extent of underreporting changes over time, then such series may not paint an accurate picture of long-run trends in top income shares. Another problem is that the income unit is either the individual or the tax filing unit, rather than the measure of income that is typically of most interest to economists: family or household income, equivalised for household size.

The issue of comparability of top incomes estimates across countries is dealt with in some detail in Atkinson (2007*a*) and Atkinson and Leigh (2007*b*). The particular focus here is on issues of comparability that affect the use of top incomes series as a panel dataset and on appropriate corrections to be made. I therefore focus on seven issues:

[7] On the construction of compatible top incomes series, see also Piketty and Saez (2006*a*).

© The Author(s). Journal compilation © Royal Economic Society 2007

the start date for the tax year, the appropriate cut-off for the adult population, the definition of the income unit, the construction of the personal income total, the definition of taxable versus total income in taxation statistics, the inclusion of capital gains and interpolation of data in missing years.[8]

1 *The tax year.* In Canada, France, Japan, the Netherlands, Spain, Sweden, Switzerland and the US, the tax year and calendar year are one and the same. However, this is not true of all countries. The tax year commences on July 1 in Australia, April 1 in New Zealand, and April 6 in Ireland and the UK. In order to construct a panel dataset of top incomes, which might be matched to data collected on a calendar-year basis, I create a dataset in which top income shares are averaged across tax years for these countries. In referring to tax changes in this article, any reference to a tax year should be taken to refer to the start of the tax year – for example, the 1980 Australian tax year is the tax year starting on July 1, 1980.

2 *The appropriate age cut-off for the adult population.* The estimates for Australia, the Netherlands, New Zealand and the UK use persons aged 15 and over, the estimates for Sweden use persons aged 16 and over, the estimates for Ireland use persons aged 18 and over, while those for Canada, France, Japan, Spain, Switzerland and the US use persons aged 20 and over. To give some sense of the magnitude of the effect, Atkinson and Leigh (2005, 2007*a*) find for Australia and New Zealand that shifting from a population control total of 15 and over to one of persons aged 20 and over reduces the top 1% share by approximately 0.5 percentage points and the top 10% share by approximately 2 percentage points. They do not discern any substantial change in this effect over time; see also Roine and Waldenström, (2006) who show a similar robustness check for Sweden. I do not make any adjustment for this, though an argument could be made for doing so.

3 *The income unit.* In Australia, Canada and Spain, the tax unit is the individual. In France, Ireland, the Netherlands, Switzerland and the US, the tax unit is a married couple or single individuals, and the population control total is therefore the adult population minus the number of married females. Germany has a hybrid system, with most taxpayers filing as tax units and the very rich filing as individuals. In 1948, the US changed the incentives for married women to file separately, so Piketty and Saez adjust the income shares by 'about 2.5%' for the period 1913–47 (Piketty and Saez, 2001, 35n). A more significant shift occurred in Japan (1950), New Zealand (1953), Sweden (1971) and in the UK (1990), when the tax unit switched from the household to the individual. In the case of Japan, Moriguchi and Saez (2006) are able to subtract dependent income from head-of-household income for earlier years. For Sweden, Roine and Waldenström (2006) find little impact of this shift, so they do not adjust their series. For New Zealand and the UK, such a correction is not possible, and

[8] For brevity, I do not deal with two other issues. On the treatment of part-year units, see Atkinson and Salverda (2003). On the use of the Pareto extrapolation method to estimate the share of top income groups where the share of the population in the top income band is larger than the share to be estimated, see Atkinson (2007*a*).

© The Author(s). Journal compilation © Royal Economic Society 2007

the effect of the switch appears to have been to increase top income shares in both countries substantially. Atkinson and Leigh (2005) therefore adjust the New Zealand series, assuming that the whole of the increase in the top shares from 1952 to 1953 represented the effect of the move from a tax unit to an individual basis, and apply this constant adjustment to 1952 and all previous years. Similarly, I assume that the UK increase in top income shares from 1989 to 1990 also represented the effect of the move from a tax unit to an individual basis, and apply this constant adjustment to the years 1908–89. (Since UK top income shares were steadily rising in the 1980s and 1990s, attributing all of the change from 1989 to 1990 to the shift in the tax unit probably underestimates the true increase in top income shares.)

4 *The personal income total.* The appropriate income control total used to derive the top income shares in each country is the sum that would have been reported were all adults to have paid tax. This figure is typically derived by starting with the national accounts and subtracting the income of the government sector, corporate sector, and non-profit sector.[9] While the accuracy of the personal income control total will doubtless vary from country to country (depending largely on the quality of the national accounts), there do not appear to be systematic differences between nations.[10]

5 *Income definition – taxable and total income.* In the earlier years, taxation statistics for several countries were tabulated by assessable income (income less deductions). In later years, this shifted to total income. In the case of Australia, New Zealand and the UK, this change has been accounted for in the production of the top incomes series. Another issue is that certain types of income are not included in taxation statistics. In the case of the US, Piketty and Saez (2001) note that non-taxable (and partially taxable) social security benefits grew as a share of personal income during the post-war decades but find that these changes had only a trivial impact on top income shares. However, differences in the definition of taxable income may have a greater impact when comparing top income shares across countries.

6 *Income definition – inclusion and exclusion of realised capital gains.* For the purposes of the analysis in this article, I present series that exclude capital gains wherever possible. For Australia, Ireland and New Zealand, series excluding capital gains are not readily available, so series for these countries include realised capital gains, to the extent that such gains were taxable.

7 *Interpolation for missing years.* In several instances, taxation statistics are unavailable. For example, income taxation statistics for New Zealand are available for 1921–2002, but were not compiled during the Depression (1931–2), World War II (1941–4), and a few later years (1961, 1974 and 1976). Where the gap is four years or less, I linearly interpolate for the missing years. However, in some cases, the gap is larger than four years. For example, the share of the richest 10% in

[9] Personal income in the national accounts is typically constructed from a variety of sources, including surveys and data on wage bills. However, as Nolan (2007) points out, in some instances total taxable income may itself be used in the construction of the national accounts personal income figure.

[10] The personal income control total is about two-thirds of GDP. This ratio appears quite similar across countries, and shows no systematic trends, either upwards or downwards.

© The Author(s). Journal compilation © Royal Economic Society 2007

the United Kingdom is missing from 1920–36, and in such instances, I do not interpolate. In the case of Switzerland, taxpayers are only required to file returns every two years, so I assign the same figure to both years. During the period 1887–98, Japanese tax returns were for overlapping three years periods, so I assign the top income estimate to the middle year.[11] And for France, top income shares for 1900–10 are based on average data for the period, so I assign the number to 1905.

Appendix Table A2 presents summary statistics, and Appendix Table A3 shows correlations between the inequality measures. Figures 1 and 2 depict the top 10% share for Anglo-Saxon and non Anglo-Saxon countries; while Figures 3 and 4 show the top 1% share for these two sets of countries (note that the top 10% share is unavailable for Japan). In all countries except Switzerland, top income shares tended to fall from the 1920s to the 1970s. Since the 1970s, top income shares in the Anglo-Saxon countries (Australia, Canada, Ireland, New Zealand, the UK and the US) have risen sharply, while shares in Japan and in the continental European countries (France, the Netherlands, Spain, Sweden and Switzerland) remained relatively stable.

Across the 13 countries, using the adjusted and interpolated series, there is a total of 761 observations for the share of the richest 10%, and 937 observations for the share of the richest 1%. This is more than five times as many observations as in the LIS, and exceeds the number of high-quality country-year observations in both the Deininger and Squire database and the WIID.[12]

2. Comparison With Other Inequality Measures

While top income shares are available over a long time horizon, are they a useful measure of inequality in a society? Measured against the axioms of inequality set out in Cowell (1995), top income shares satisfy three basic principles: income scale independence, principle of population, and anonymity.[13] However, top income shares only weakly satisfy the Pigou-Dalton transfer principle, since a transfer from rich to poor will never increase the top income shares but, if the transfer is between two individuals who are both within the top group or both outside the top group, then the share measure will remain unchanged. (Top income shares are also not decomposable into within-group inequality and between-group inequality.) Another issue is that top income shares are based on pre-tax incomes. To the extent that the redistributive effect of taxation differs across countries and over time, top

[11] As Moriguchi and Saez (2006) point out, the effect of tax averaging over multiple years is probably also to reduce top income shares. Neither they nor I make any adjustment for this.

[12] Deininger and Squire identify 693 observations which they label 'accept'. Version 2a of the WIID contains 1,223 observations classified as Quality = 1, but many of these are repeated observations for the same country-year, so there are only 540 high-quality country-year observations in the WIID.

[13] Income scale independence requires that the inequality measure be unaffected by proportional changes in income (eg. expressing income in pence rather than pounds should not change inequality). The principle of population requires that the inequality measure be unaffected by replications of the population (eg. merging two identical distributions should not change inequality). Anonymity requires that the inequality measure be unaffected by characteristics apart from income.

© The Author(s). Journal compilation © Royal Economic Society 2007

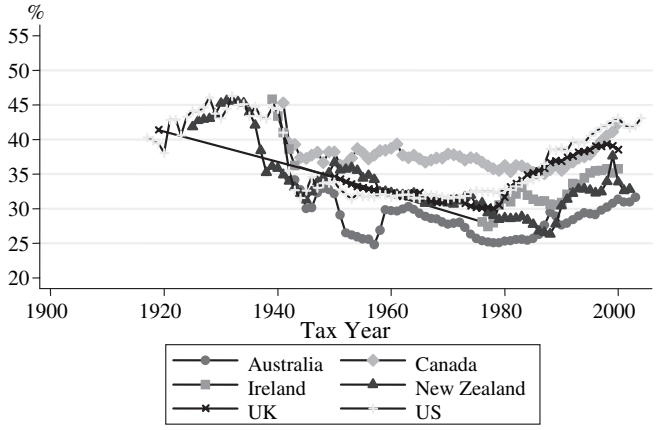

Fig. 1. *Income Share of Richest 10% in Anglo-Saxon Countries*

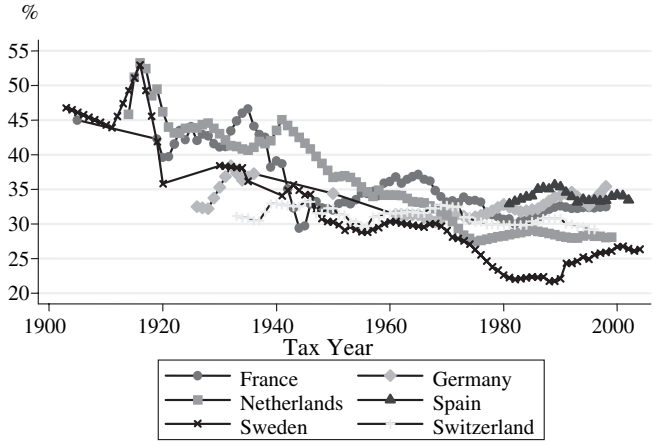

Fig. 2. *Income Share of Richest 10% in Non Anglo-Saxon Countries*

income shares may be a poor proxy for the differences in spending power in a given society.

Nonetheless, if the taxation system does not change, then a shock to the income distribution (eg. skill-biased technological change) may affect both the bottom and top of the distribution. In this event, it may be the case that the share of income held by the top 10% is a usable proxy for inequality across the distribution. One way to test whether top income shares are a good proxy for inequality across the distribution is to analyse the relationship between top incomes measures and income inequality empirically in the recent era (when both are available). In this Section, I first compare top income shares with Gini coefficients from the WIID (since the Deininger and Squire database is fully contained within the WIID, I do not separately analyse that dataset) and then compare top income shares with income measures from the LIS.

© The Author(s). Journal compilation © Royal Economic Society 2007

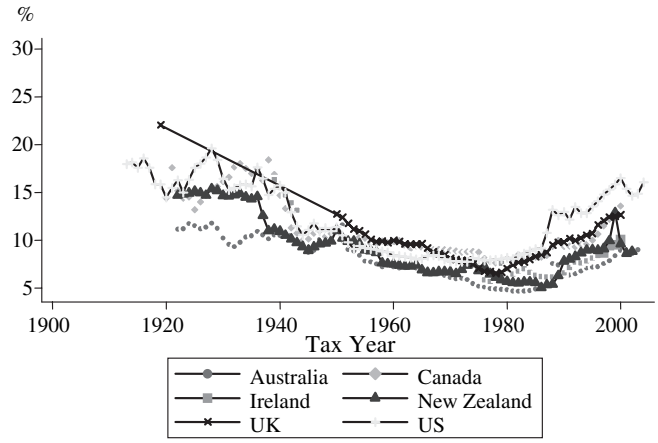

Fig. 3. *Income Share of Richest 1% in Anglo-Saxon Countries*

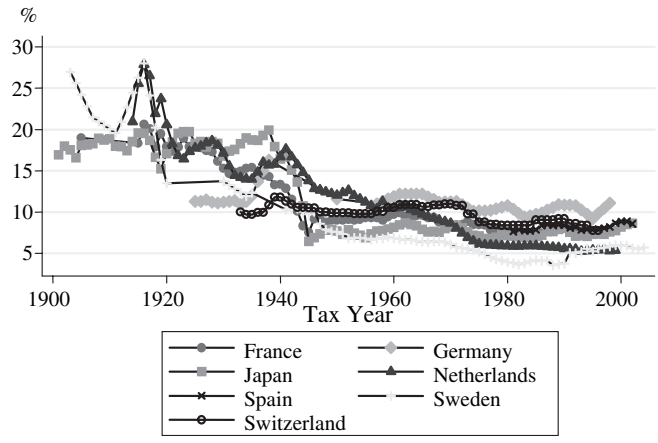

Fig. 4. *Income Share of Richest 1% in Non Anglo-Saxon Countries*

In order to analyse the relationship between top income shares and Gini coefficient in the WIID, I use observations from the WIID that meet four criteria:

(*a*) the estimate was for income rather than consumption or expenditure;
(*b*) the income-sharing unit was the family or household;
(*c*) the estimate covered the full geographic area of the country;
(*d*) the estimate covered the entire population.

Where there were multiple observations that met these standards, I used the observation given the highest quality rating by the WIID.

To see the relationship between top income shares and other measure of inequality, I simply regress one upon the other. In principle, it does not matter which is the dependent variable but here I use the top income share as the dependent variable, since it then becomes straightforward to extend the model to estimate

© The Author(s). Journal compilation © Royal Economic Society 2007

specifications with more than one inequality measure on the right-hand side of the equation.

The estimating equations take the following form:

$$\text{Log}(S)_{jt} = \alpha + \beta \text{Log}(\textit{Ineq measure})_{jt} + \varepsilon_{jt} \quad (1)$$

$$\text{Log}(S)_{jt} = \alpha + \beta \text{Log}(\textit{Ineq measure})_{jt} + \gamma_j + \varepsilon_{jt} \quad (2)$$

$$\text{Log}(S)_{jt} = \alpha + \beta \text{Log}(\textit{Ineq measure})_{jt} + \gamma_j + \delta_t + \varepsilon_{jt}, \quad (3)$$

where S is a measure of top income (such as the income share of the top 10%) in country j in year t, and *Ineq measure* is some alternative measure of inequality. Equation (2) also includes a country-specific term, γ. Equation (3) is a standard panel data specification, including both country fixed effects and year fixed effects, δ.

Table 1 shows the results of this estimation. In Panel (*a*), I estimate the relationship between the top 10% share and the WIID Gini coefficient: without fixed effects; with country fixed effects; and with country and year fixed effects. The two series are positively associated with one another, with the relationship being significant at the 1% level. In Panel (*b*), I use the top 1% share as the dependent variable, and again find a positive and statistically significant relationship with the WIID Gini coefficient.

Next, I investigate the relationship between top incomes measures and the LIS. These measures of inequality are derived by the LIS team from national survey microdata, and are standardised according to the following five rules:

(*a*) the income measure is disposable income;
(*b*) income is pooled within households and divided by the square root of the number of people in the family;
(*c*) all individuals, including children, are weighted according to their representation in the population;
(*d*) income is bottom-coded at 1% of equivalised mean income, and top-coded at 10 times mean income; and
(*e*) missing and zero incomes are excluded.[14]

The LIS provides a number of inequality measures. Here, I use the Gini, the Atkinson index (with an inequality aversion parameter of 1), the 90:10 ratio, the 90:50 ratio and the 50:10 ratio. Since Japan and New Zealand are not included in the LIS, the regressions below cover only 11 countries.

Table 2 shows the results from this exercise. Without country and year fixed effects (Panel (*a*)), the top 10% share is positively related to each of the other inequality measures, with the relationship being statistically significant at the 1% level. Somewhat surprisingly, the 50:10 ratio is significantly related to the top 10% share, and this relationship remains significant even holding constant the 90:50 ratio. Results including country fixed effects (Panel (*b*)) are similar to those without country fixed

[14] For more detail on the methodology used to construct the LIS inequality measures, see http://www.lisproject.org/keyfigures/methods.htm.

© The Author(s). Journal compilation © Royal Economic Society 2007

Table 1

Top Incomes and WIID Inequality Measures

(*a*) *Dependent variable is Ln(Top 10% Share)*			
	(1)	(2)	(3)
Ln(Gini)	0.304***	0.229***	0.219***
	[0.045]	[0.042]	[0.038]
Country FE	No	Yes	Yes
Year FE	No	No	Yes
Observations	263	263	263
R-squared	0.20	0.76	0.89
(*b*) *Dependent variable is Ln(Top 1% Share)*			
Ln(Gini)	0.799***	0.693***	0.422***
	[0.086]	[0.100]	[0.070]
Country FE	No	Yes	Yes
Year FE	No	No	Yes
Observations	300	300	300
R-squared	0.29	0.67	0.89

Note. Robust standard errors in brackets. *, ** and *** denote significance at the 10%, 5% and 1% levels respectively.

effects, except that the relationship between the 50:10 and the top 10% share is insignificant once the 90:50 ratio is included in the regression.

Panel (*c*) includes both country and year fixed effects, allowing for country-specific differences in the relationship between the inequality measures, as well as for non-linear time variation. In this specification, the Gini, Atkinson index and 90:50 ratio are each positively and significantly associated with the share of the richest 10%, while the 50:10 ratio is negatively and significantly related to the share of the richest 10%. Again, when both the 90:50 and the 50:10 ratios are included in the regression, only the 90:50 ratio is statistically significant.

Table 3 replicates the exercise, using the top 1% share as the dependent variable. The results are similar to those in the previous Table, with each of the inequality measures being positively and significantly related to the income share of the richest 1% (Panel (*a*)). This remains true (with the exception of the 50:10 ratio) when country fixed effects are added to the regression (Panel (*b*)). Including country and year fixed effects (Panel (*c*)), the coefficients are mostly positive but only statistically significant for the 90:50 ratio (which is positively related to the share of the richest 1%), and the 50:10 ratio (which is negatively related to the share of the richest 1%).

Using inequality data from either the WIID or the LIS, it appears that the relationship between top income shares and other inequality measures remains strong even when country fixed effects are included. This suggests that within-country *changes* in top income shares can be a useful proxy for *changes* in other inequality measures. Indeed, the relationship between the top 10% share and several other inequality measures remains statistically significant even with both country and year fixed effects.

3. Conclusion

The careful creation of top incomes series over recent years provides a window into the long-run distribution of incomes in an (increasing) number of nations. But using these

© The Author(s). Journal compilation © Royal Economic Society 2007

Table 2

Top Incomes and LIS Inequality Measures

Dependent variable: Ln(Top 10% Share)						
(*a*) *Without Fixed Effects*						
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Gini)	0.824***					
	[0.072]					
Ln(Atkinson Index $\varepsilon = 1.0$)		0.391***				
		[0.058]				
Ln(90:10)			0.513***			
			[0.060]			
Ln(90:50)				1.203***		0.874***
				[0.135]		[0.178]
Ln(50:10)					0.662***	0.311***
					[0.101]	[0.098]
Observations	63	63	63	63	63	63
R-squared	0.56	0.42	0.52	0.5	0.39	0.55
(*b*) *With Country Fixed Effects*						
Ln(Gini)	0.881***					
	[0.098]					
Ln(Atkinson Index $\varepsilon = 1.0$)		0.306***				
		[0.060]				
Ln(90:10)			0.534***			
			[0.113]			
Ln(90:50)				1.126***		1.082***
				[0.182]		[0.184]
Ln(50:10)					0.339**	0.178
					[0.154]	[0.147]
Observations	63	63	63	63	63	63
R-squared	0.91	0.86	0.85	0.88	0.8	0.89
(*c*) *With Country and Year Fixed Effects*						
Ln(Gini)	0.445**					
	[0.171]					
Ln(Atkinson Index $\varepsilon = 1.0$)		0.140*				
		[0.080]				
Ln(90:10)			0.137			
			[0.171]			
Ln(90:50)				0.707***		0.674***
				[0.215]		[0.239]
Ln(50:10)					−0.227*	−0.123
					[0.133]	[0.135]
Observations	63	63	63	63	63	63
R-squared	0.98	0.98	0.98	0.99	0.98	0.99

Notes. Robust standard errors in brackets. *, ** and *** denote significance at the 10%, 5% and 1% levels respectively.

data as a long panel requires careful attention to the various differences between them. This article highlights the main disparities between the series and, where possible, makes adjustments to account for these. Such data will not be perfectly comparable but such is the nature of many of the existing datasets used to measure the causes and effects of inequality across countries.

The other question that this article has sought to answer is whether top incomes series are a useful proxy for inequality across the income distribution. On a theoretical level, this seems plausible, since many of the factors that affect inequality are likely to

© The Author(s). Journal compilation © Royal Economic Society 2007

Table 3

Top Incomes and LIS Inequality Measures

Dependent variable: Ln(Top 1% Share)						
(*a*) *Without Fixed Effects*						
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Gini)	1.495***					
	[0.203]					
Ln(Atkinson Index $\varepsilon = 1.0$)		0.688***				
		[0.142]				
Ln(90:10)			0.882***			
			[0.154]			
Ln(90:50)				2.155***		1.714***
				[0.346]		[0.415]
Ln(50:10)					1.106***	0.418*
					[0.248]	[0.250]
Observations	63	63	63	63	63	63
R-squared	0.44	0.31	0.37	0.38	0.27	0.41
(*b*) *With Country Fixed Effects*						
Ln(Gini)	1.919***					
	[0.298]					
Ln(Atkinson Index $\varepsilon = 1.0$)		0.615***				
		[0.157]				
Ln(90:10)			1.017***			
			[0.292]			
Ln(90:50)				2.407***		2.373***
				[0.497]		[0.554]
Ln(50:10)					0.489	0.137
					[0.370]	[0.456]
Observations	63	63	63	63	63	63
R-squared	0.83	0.76	0.75	0.8	0.7	0.8
(*c*) *With Country and Year Fixed Effects*						
Ln(Gini)	0.797					
	[0.602]					
Ln(Atkinson Index $\varepsilon = 1.0$)		0.165				
		[0.262]				
Ln(90:10)			0.178			
			[0.474]			
Ln(90:50)				1.483*		1.373
				[0.786]		[0.843]
Ln(50:10)					−0.614*	−0.401
					[0.355]	[0.439]
Observations	63	63	63	63	63	63
R-squared	0.96	0.95	0.95	0.97	0.96	0.97

Notes. Robust standard errors in brackets. *, ** and *** denote significance at the 10%, 5% and 1% levels respectively.

have an impact on both the top and bottom of the distribution. Comparing measures of inequality based on top income shares with measures of household or family inequality from the WIID and LIS, I find a strong positive relationship between the series, which is robust to the inclusion of country and year fixed effects. This should be reassuring for potential users of top income shares as a proxy for inequality across the distribution, since the inclusion of country and year fixed effects is standard in cross-country panel data analysis. In summary, top income shares are far from perfect as a measure the distribution of income across society. But where other data sources are limited, they may help to fill in some of the gaps.

© The Author(s). Journal compilation © Royal Economic Society 2007

Appendix

Table A1

Sources and Adjustments

Country	Source	Adjustments	Years Covered (Adjusted Series)
Australia	Atkinson and Leigh (2007*a*, Table 1)	Converted to calendar year basis.	1922–2003
Canada	Saez and Veall (2005, Excel Table B1)	No adjustments made.	1920–2000
France	Piketty (2007, Table 13.1)	Top income shares for 1900–10 are based on average data for the period, so this number is assigned to 1905.	1905–1998
Germany	Dell (2007, Table 13.7)	No adjustments made.	1925–1998
Ireland	Nolan (2007, Table 13.10)	Converted to calendar year basis.	1939–2000
Japan	Moriguchi and Saez (2006, Table A1)	No adjustments made. No top 10% series available.	1886–2002
Netherlands	Salverda and Atkinson (2007, Table 13.8)	No adjustments made.	1914–1999
New Zealand	Atkinson and Leigh (2005, Tables 1 and 3)	Adjusted 1% series taken from Table 3. Unadjusted 10% series taken from Table 1 and adjusted in a similar manner. Both series then converted to calendar year basis.	1922–2002
Spain	Alvaredo and Saez (2006, Table B2)	No adjustments made	1981–2002
Sweden	Roine and Waldenström (2006, Excel Table A2)	No adjustments made.	1903–2004
Switzerland	Dell *et al.* (2007, Table 13.9)	Taxpayers are only required to file returns every two years, so I assign the same figure to both years.	1933–1996
United Kingdom	Atkinson (2007*b*, Table 13.2)	In 1908–89, 10% share multiplied by 1.081 and 1% share multiplied by 1.130, to take account of the shift from joint to individual filing in 1990. Converted to calendar year basis.	1919–2000
United States	Piketty and Saez (2006*b*, Excel Table A1)	No adjustments made.	1913–2004

Note. For all countries, top income shares in missing years are linearly interpolated, so long as the gap is four years or less.

Table A2

Summary Statistics

Variable	Obs	Mean	SD	Min	Max
Top 10% share	761	34.103	5.556	21.700	53.310
Top 1% share	937	10.634	4.258	3.570	28.040
Gini (WIID)	300	33.171	6.816	19.100	54.300
Gini (LIS)	63	29.317	3.872	19.700	37.200
Atkinson Index					
ε=1.0 (LIS)	63	15.348	3.683	7.300	22.900
90:10 (LIS)	63	3.886	0.828	2.430	5.850
50:10 (LIS)	63	2.052	0.302	1.581	2.799
90:50 (LIS)	63	1.878	0.163	1.510	2.230

Note. Sample is those country-year for which either the top 1% share or the top 10% share is available.

© The Author(s). Journal compilation © Royal Economic Society 2007

Table A3

Correlation Coefficients

	Top 10%	Top 1%	Gini (WIID)	Gini (LIS)	Atkinson	90:10	50:10	90:50
Top 10%	1.000							
Top 1%	0.897	1.000						
Gini (WIID)	0.370	0.486	1.000					
Gini (LIS)	0.726	0.631	0.684	1.000				
Atkinson	0.622	0.541	0.714	0.934	1.000			
90:10	0.698	0.592	0.648	0.912	0.872	1.000		
50:10	0.627	0.519	0.585	0.773	0.768	0.953	1.000	
90:50	0.674	0.569	0.618	0.952	0.851	0.837	0.637	1.000

Note. All correlations are statistically significant at the 1% level.

Australian National University

References

Alvaredo, F. and Saez, E. (2006). 'Income and wealth concentration in Spain in a historical and fiscal perspective', CEPR Discussion Paper 5836.

Atkinson, A.B. (2005). 'Top incomes in the UK over the twentieth century', *Journal of the Royal Statistical Society, Series A*, vol. 168 (February), pp. 325–43.

Atkinson, A.B. (2007*a*). 'Measuring top incomes: methodological issues', in Atkinson and Piketty, (2007), pp. 18–42.

Atkinson, A.B. (2007*b*). 'Top incomes in the United Kingdom over the twentieth century', in Atkinson and Piketty (2007), pp. 82–140.

Atkinson, A.B. and Brandolini, A. (2001). 'Promise and pitfalls in the use of "secondary" data-sets: income inequality in OECD countries as a case study', *Journal of Economic Literature*, vol. 39 (September), pp. 771–99.

Atkinson, A.B. and Leigh, A. (2005). 'The distribution of top incomes in New Zealand', Australian National University CEPR Discussion Paper 503.

Atkinson, A.B. and Leigh, A. (2007*a*). 'The distribution of top incomes in Australia', *Economic Record*, vol. 83(262), pp. 247–61.

Atkinson, A.B. and Leigh, A. (2007*b*). 'The distribution of top incomes in five Anglo-Saxon countries over the twentieth century', mimeo, Australian National University.

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

Atkinson, A.B. and Salverda, W. (2003). 'Top incomes in the Netherlands and the United Kingdom over the twentieth century', discussion paper.

Banerjee, A. and Piketty, T. (2005). 'Top Indian incomes, 1922–2000', *The World Bank Economic Review*, vol. 19, (December), pp. 1–20.

Cowell, F.A. (1995). *Measuring Inequality* (2nd edition), Hemel Hempstead: Harvester Wheatsheaf.

Deininger, K. and Squire, L. (1996). 'A new data set measuring income inequality,' *World Bank Economic Review*, vol. 10, (September), pp. 565–91.

Dell, F. (2005). 'Top incomes in Germany and Switzerland over the twentieth century', *Journal of the European Economic Association*, vol. 3(2–3), (April/May), pp. 412–21.

Dell, F. (2007). 'Top incomes in Germany throughout the twentieth century: 1891–1998', in Atkinson and Piketty (2007), pp. 365–425.

Dell, F., Piketty T. and Saez, E. (2007). 'Income and wealth concentration in Switzerland over the 20th century', in Atkinson and Piketty (2007), pp. 472–500.

Fearon, J. and Laitin, D. (2003). 'Ethnicity, insurgency, and civil war', *American Political Science Review*, vol. 97(1), (February), pp. 75–90.

Gilbert, G. (1997). 'Adam Smith on the nature and causes of poverty', *Review of Social Economy*, vol. 55(3), pp. 273–91.

Kuznets, S. (1953). *Shares of Upper Income Groups in Income and Savings*, New York: National Bureau of Economic Research.

© The Author(s). Journal compilation © Royal Economic Society 2007

Kuznets, S. (1955). 'Economic growth and income inequality', *American Economic Review*, vol. 65, (March), pp. 1–28.
Leigh, A. and van der Eng, P. (2007). 'Top incomes in Indonesia, 1920–2004', Australian National University CEPR Discussion Paper 549.
Moriguchi, C. and Saez, E. (2006). 'The evolution of income concentration in Japan, 1885–2002: evidence from income tax statistics', National Bureau of Economic Research Working Paper 12558.
Nolan, B. (2007). 'Long-term trends in top income shares in Ireland', in Atkinson and Piketty, (2007), pp. 501–30.
Piketty, T. (2001). *Les Hauts Revenus en France au 20ème siècle*. Paris: Grasset.
Piketty, T. (2003). 'Income inequality in France, 1901–1998', *Journal of Political Economy*, vol. 111(5), (October), pp. 1004–42.
Piketty, T. (2007). 'Income, wage and wealth inequality in France, 1901–1998', in Atkinson and Piketty, (2007), pp. 43–81.
Piketty, T. and Qian, N. (2006). 'Income inequality and progressive income taxation in China and India, 1986–2015', CEPR Discussion Paper 5703.
Piketty, T. and Saez, E, (2001). 'Income inequality in the United States, 1913–1998', National Bureau of Economic Research Working Paper 8467.
Piketty, T. and Saez, E. (2003). 'Income inequality in the United States, 1913–1998', *Quarterly Journal of Economics*, vol. 118(1), (February), pp. 1–39.
Piketty, T. and Saez, E, (2006*a*). 'The evolution of top incomes: a historical and international perspective', *American Economic Review*, vol. 96(2), pp. 200–5.
Piketty, T. and Saez, E. (2006*b*). 'Income inequality in the United States', Tables and Figures updated to 2004 in Excel format, http://emlab.berkeley.edu/users/saez/ (downloaded 6 December 2006).
Roine, J. and Waldenström, D. (2006). 'Top incomes in Sweden over the twentieth century', Research Institute of Industrial Economics Working Paper 667, Stockholm, Sweden.
Saez, E. and Veall, M. (2005). 'The evolution of high incomes in Northern America: lessons from Canadian evidence', *American Economic Review*, vol. 95(3), (June), pp. 831–49.
Salverda, W. and Atkinson, A. B. (2007). 'Top incomes in the Netherlands over the twentieth century', in Atkinson and Piketty, (2007), pp. 426–72.

© The Author(s). Journal compilation © Royal Economic Society 2007

Copyright of Economic Journal is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.