

The Economic and Social Review, Vol. 37, No. 2, Summer/Autumn, 2006, pp. 149-173

Mobility and Gender at the Top Tail of the Earnings Distribution*

ROSS FINNIE

Statistics Canada

IAN IRVINE

Concordia University

Abstract: The increasing share of the top fractile in the earnings distributions of several Anglo-Saxon heritage economies since the 1970s has been dramatic, and well documented. To date, however, little is known about the socio-economic origins and gender composition of the very top tail in the modern era. This paper takes a first step in filling some of the holes in our knowledge. We use a tax-filer data base for Canada for the period 1983-2003 that contains about eighty million observations. We show first that male earners in the top one thousandth of the distribution come very disproportionately from families with incomes in the top decile. In contrast, individuals in the remaining part of the top centile have more dispersed socio-economic origins. Second we show that female participation in the top fractiles has been very low, and that growth in participation has been slow yet definite. In contrast, female earnings in this echelon are almost on par with male earnings. Third, we show that there is an enormous asymmetry between the genders when it comes to spousal earnings: high-earning women have very high-earning spouses, but not vice versa. 'Secondary males' have earnings levels almost ten times as high as 'secondary females', suggesting that, even at this extremely elevated earnings level there is truth to the adage about who lies 'behind' successful individuals. Finally, it is illustrated that the earnings concentration that has characterised the last three decades did not change with the end of the 'tech boom' in the year 2000.

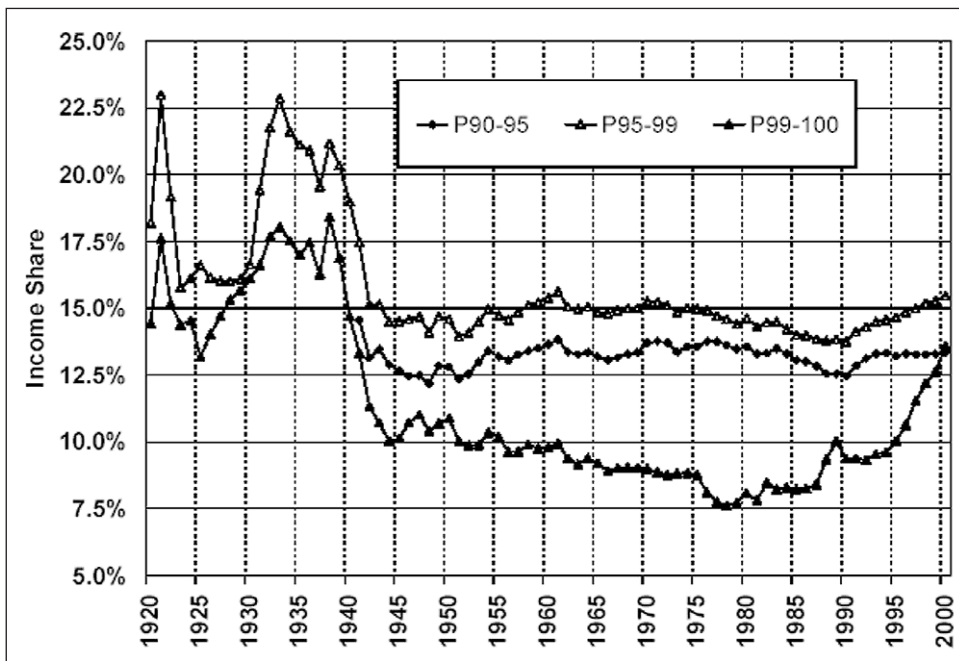
*An earlier version was presented at the Conference on *Macroeconomic Perspectives* in Honour of Brendan M. Walsh, held at University College Dublin on 7 October, 2005. Ross Finnie works at Statistics Canada, and is a Fellow of the School of Policy Studies, Queen's University, Ontario, Canada. Ian Irvine is a Professor at Concordia University, Montreal, Quebec, Canada, H4X 1T9. The authors are grateful to Statistics Canada for providing access to the data base, to Andre Bernard for excellent research assistance, and to the Journal editor and referees for comments on the early drafts of this paper. Corresponding author is Irvine at irvinei@vax2.concordia.ca.

I INTRODUCTION

The recent growth in share of the very top fractiles of the income distribution in Anglo-Saxon heritage economies has been truly remarkable. Saez and Veall (2005) show that the pattern of gain has been broadly similar in the United States and Canada, with the latter lagging slightly: the wage-income share of the top 1 per cent has approximately doubled from 6 per cent to 12 per cent over the period 1975 to 2000, and the share of the top 0.1 per cent has approximately quadrupled over the same time frame – from just over 1 per cent to 5 per cent. These patterns are illustrated in Figures 1 and 2, which are taken from Saez and Veall.

Atkinson (2005) and Atkinson and Leigh (2004) have shown that a similar pattern characterises the data for the UK, Australia and New Zealand, though the trends in the Antipodes have not been quite so strong. A recent paper for Ireland – another primarily English-speaking economy – by Nolan (2005) indicates a pattern of top fractile growth similar to the economies examined by Atkinson and Leigh. Piketty and Saez (2003) initially charted the patterns for

Figure 1: *Top Decile Shares, Canada*



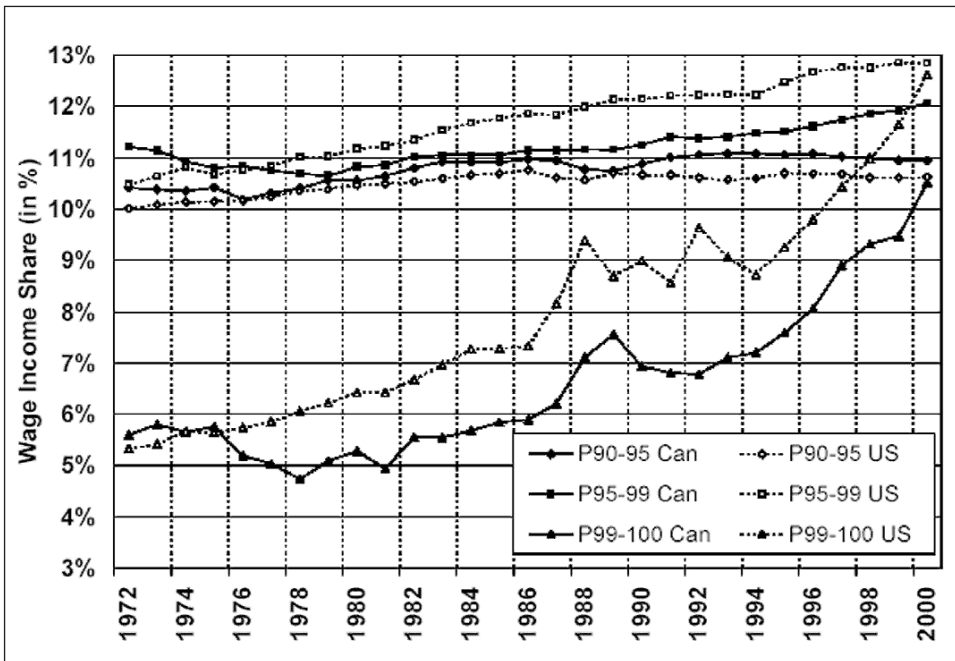
Source: Saez and Veall (2005).

the US. In contrast, many other studies have found no such pattern in the data for France (Piketty, 2003), Japan (Moriguchi and Saez, 2005), the Netherlands (Atkinson and Salverda, 2005), Switzerland (Dell, Piketty and Saez, 2005) or Spain (Alvaredo and Saez, 2005) as examples.

Such a growth in concentration raises a number of questions about those individuals who inhabit the very top echelon of the distribution: what is the degree of social mobility in this world? Does there exist a ‘glass ceiling’ for women that restricts their participation in this group? For those women who are members of this elite, are their earnings comparable to male earnings? What are the spousal earnings of individuals in the top fractiles? How have these patterns evolved in recent decades? To date the literature has shed very little light on these issues.

These questions are of interest not just to economists, but to sociologists and public policy makers broadly defined. In the present paper we explore these questions and succeed in providing answers with the help of a remarkable data base composed of about 80 million Canadian tax filers. While we obviously cannot generalise our findings to other economies, even English-

Figure 2: *Top Decile Shares, Canada and the US*



Source: Saez and Veall (2005).

speaking ones, our results will hopefully motivate similar investigations elsewhere.

But first, why are we interested? For some, the concentration of income and/or wealth may be just the inevitable consequence of the incentives necessary for a more efficient functioning of a market economy. If this is so, then we could take the position that, provided the poor are not completely destitute, there is little reason for concern: the growth in the top tail incomes may not have been at the expense of those at the bottom in the distribution and, therefore, any resulting increase in inequality or polarisation may not reflect a worsening of the absolute status of the poor.

An alternative perspective sees polarisation as potentially welfare reducing, even if it is brought about by an increase in the distance of the top tail from the median without increasing the distance between the median and those in the lower tail. Increased polarisation may result in elevated crime rates or even social disintegration in the extreme (Diamond, 2005).¹ Crime may also spread in less equal societies (Fajnzylber, Lederman and Loayza, 2002) because the opportunity costs to the disenfranchised fall. Rioting by marginalised groups in France in the Autumn of 2005 was arguably a reflection of the divide between insiders and outsiders.²

A second consequence of increasing income spreads is reduced social mobility. Higher incomes in the upper centiles reduce the likelihood of intergenerational income mobility as a result of bequests and trusts. Recent evidence (Corak, 2004) suggests that social mobility in the US may now be as low as some of 'Old Europe', although mobility in Canada is high, despite the recent surge in top incomes there. Reduced mobility may also result from a more easily acquired and financed education on the part of the offspring of those with very high incomes.

Third, increases in income concentration at the top tail cannot but lead to increases in the concentration of wealth. Kennickell (2001), indicates that the top 1 per cent of the US wealth distribution now owns about 35 per cent of all personal wealth, the next 9 per cent, owns about one-third, and the remaining 90 per cent of the population owns less than one-third. Concentrations of wealth also lead to reduced social mobility, and an increase in political power. Glaeser (2005) proposes that greater concentrations of wealth can increase the lobbying and political power of those at the upper tail and thereby increase inequality further.

¹As an example, *The Washington Post* reported in September 2005 that, in Brazil, 36,000 people had been killed with guns in 2004. Sala-i-Martin (2002) indicates that Brazil's income distribution not only displays much more inequality than virtually all countries at a comparable level of development, but that it is actually bimodal.

²Duclos, Esteban and Ray (2004) examine the link between alienation and polarisation.

Fourth, greater polarisation may decrease the demand for common public goods. The rich may prefer gated communities built around golf courses with their own security forces, road maintenance, and local public goods, while the greater part of the income distribution may desire more public-school investment. A substantial literature (e.g., Alesina, Baqir and Easterly, 1999) exists on this, that shows not only does the polarisation of demands, itself driven by income polarisation, decrease redistributive possibilities, but can also lead to inefficient levels of human capital investment for the economy at large.

Finally there is the question of public policy. If it is desirable to place some limit on polarisation, or income/wealth concentration, it is necessary to understand why and how the observed concentrations develop and change. Are the giant shifts in shares recently observed the result of taxation policy, or changes in social norms, or the fruits of invention, technological shocks and luck? If public policy measures have indeed been behind these seismic shifts, then we need to understand the linkages between public policy and wealth/income concentrations.

Understanding why earnings have become so concentrated at the top tail is additionally important in view of the role that has been attributed to technological change in the last decade. The accepted wisdom is that technological change has been of the skill-biased type, and hence benefited those with high skills and incomes. At the same time, the fact that only the top 1 per cent of the distribution has increased its share of total earnings in Canada (and several of the English-speaking economies) in a substantial way, whereas one would expect that the whole of the top decile at least would have benefited from skill-biased technological change, raises the possibility that the benefits of skill-biased technological change have been largely appropriated by a very small group. Eckstein and Nagypal (2004) indicate that the returns to education have been very non-linear, and that postgraduates, rather than university graduates at large, have gained disproportionately in the last three decades in the US. Postgraduates are very small in number and therefore this pattern is consistent with what emerges from an examination of the very top tail of the earnings distribution.

Saez and Veall (2005) and Piketty and Saez (2003) have each proposed that the growing share of the top fractiles can be understood best as reflecting primarily a change in social norms – at least in Anglo-Saxon heritage economies, rather than being attributable to a change in tax structure (as suggested by Feldstein (1995)), the impact of skill-biased technological change, or the spectacular growth of stock options in the nineties.

The current paper examines the characteristics of those individuals who compose the top echelons of the distribution. In line with recent practice, we

compare the top one thousandth of the distribution with the remainder of the top centile and in turn with the remainder of the top decile and the rest of the population. We first focus on intergenerational mobility. Buchinsky and Hunt (1999) have found that income mobility has lessened substantially in the US in recent decades. They used a representative sample from the NLSY. In contrast we investigate the extreme top tail of the distribution – though our data do not permit us to examine how mobility has changed over time – for reasons that will become clear in our discussion of the data. Second, we examine the composition of the top fractiles from a gender standpoint. There exists an extensive literature on women in the labour market, and a subset of this focuses upon what is termed the ‘glass ceiling’ – proposing that there exists a less visible set of obstacles to the progress of women into the very top salary ranks. As part of this search we examine the earnings of spouses of those in the top fractiles. In each case the novelty of the present work springs from the fact that we examine the extreme top of the income distribution, a focus made possible by our access to tax-based data that accurately captures this tail.

II DATA

The Longitudinal Administrative Database (LAD) represents a 20 per cent longitudinal sample of Canadian tax-filers constructed from Canada Customs and Revenue Agency (CCRA – previously Revenue Canada) tax files. The first year of data is 1983 and the LAD currently goes to 2003, thus determining the period covered by this analysis. The relatively recent final year of the data allows us to examine patterns in the post tech-bubble period. The LAD file possesses on the order of five million observations per recent year, and the work reported here is based on the full LAD base, making for generous samples at the extremes of the distribution.

Individuals are selected into the LAD according to a random number generator based on Social Insurance Numbers (SINs) and are followed over the time by the same means.³ Individuals drop out of the LAD if they do not file taxes, or if they die or emigrate, while new filers are continually added to the database in the fixed 20 per cent of population ratio. The size of the file has grown over time, commensurate with the growth in the underlying tax-filing population.

³Individuals are tracked across any changes in their SIN, an event that characterises approximately 3 per cent of all tax filers in any given year.

The LAD's coverage of the adult population is good since the rate of tax filing is very high. Higher- and middle-income Canadians are required by law to do so, while lower income individuals have incentives to file in order to recover income tax and other payroll tax deductions made through the year and, since 1986, to benefit from various tax credits and the National Child Tax Benefit introduced in the nineties. The full set of tax files from which the LAD is constructed are estimated to cover upwards of 90 per cent of the target population, thus comparing favourably with other databases. Furthermore, a high proportion of those not included are non-labour market participants (older married women in particular), and it is on account of this status that they do not file taxes.⁴ But from the standpoint of the present study, the LAD has a very comprehensive coverage of those at the top end of the income distribution. Moreover, the work of Saez and Veall is partly based on this data set, and therefore our estimates are comparable to theirs.

The LAD includes information derived from individuals' tax files: basic demographic characteristics, income (sources, amounts), deductions, and tax paid.⁵ Important to this analysis is the quality of the income information: the income variable is constructed in a consistent manner over the full period we analyse.⁶

When sorting individuals into family units, the main criterion is the address of the individual(s). This means that both common law and standard (legal) marriages are recognised. The resulting household unit therefore corresponds closely to what is usually termed an 'economic' family – one that shares resources by virtue of living at the same domicile. The data are also characterised by a constant definition of the filing unit, unlike many other economies' data bases, where the filing unit has changed from the family to the individual, or gives the option of filing in either status (see, for example, the discussion in Atkinson and Leigh (2004)).

III SAMPLE SELECTION

Most of the samples in our analysis are constructed by imposing the following restrictions on a year-by-year (cross-sectional) basis. First, individuals had to be between 20 and 59 years of age in a given year, thus eliminating

⁴See Atkinson *et al.* (1992) for discussion of the general advantages of administrative data of this type over survey data in terms of sample representation, the accuracy of the information collected, and in other ways.

⁵For those matched into families, information pertaining to the other family members and the relevant family totals are appended to individuals' records.

⁶For most of its surveys, Statistics Canada now attempts to gain participants' permission to use their tax files to gather data on incomes.

many students and semi-retired workers whose earnings tend to be low and to vary from year-to-year for their own special reasons. Second, full-time post-secondary students were excluded (based on various tuition and other education-related deductions) on the grounds that they have a diminished attachment to the labour force and their earnings do not necessarily reflect labour-market opportunities. The results of applying these selection criteria to the 20 per cent of population LAD yielded samples for the top one thousandth of the distribution of between two and three thousand individuals in each year.

3.1 *Social Mobility*

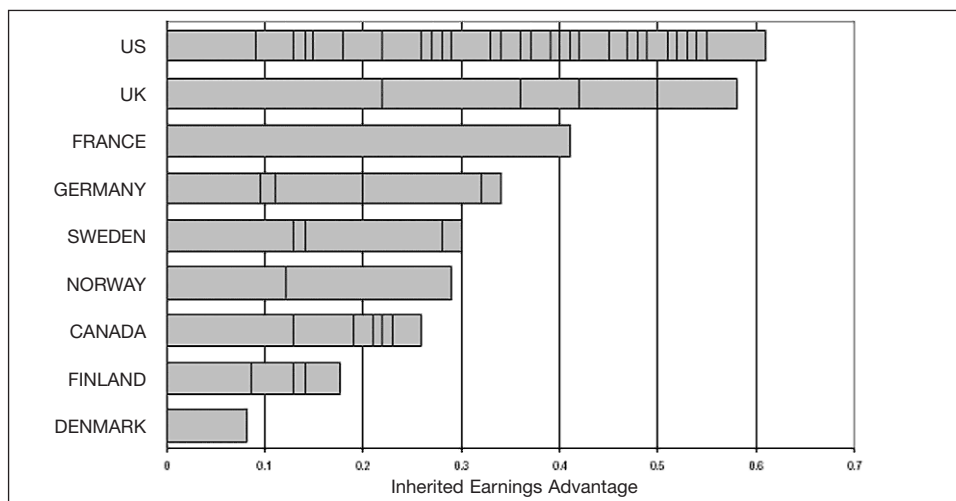
Economic and social immobility between generations is a particular consequence of economic inequality. Societies may be more willing to tolerate inequalities and extremes if such mobility is high rather than low. North American economies were traditionally thought of as being more open and mobile than the 'old world' economies of nineteenth-century Western Europe. But that appears to have changed in the twentieth century. Scandinavian economies not only have lower rates of poverty and inequality than the US and Canada but have higher rates of intergenerational economic mobility than the US. In addition, the Canadian and US economies have diverged where intergenerational mobility is concerned.

Figure 3 below is from Corak (2004). It plots various intergenerational elasticity values relating mostly the earnings of fathers and sons. Among the three English-speaking economies, the similarity of top tail share growth breaks down when it comes to intergenerational mobility. Elasticities for the US and UK are substantially higher than for Canada, whose estimates lie in the 'Scandinavian range'. The technical details of this literature were first mapped out by Solon (1989) and (1992). Corak emphasises that methodological differences explain much of the variability in the US results and, therefore, care must be exercised in cross-country comparisons.

While the *Forbes Four Hundred* yields socio-economic information on a small number of very top wealth holders, to our knowledge no work has yet been undertaken that explicitly focuses upon the socio-economic origins of those in the very top tail of the distribution. Since the LAD begins in the early eighties it affords the potential to explore the backgrounds of relatively young individuals currently in the top tail: tracing the parental income stratum of current high earners involves locating these earners in their parental family while they were still members of that household, which is when they were in their late teens. Accordingly, our strategy is to take all individuals in the LAD falling in the age bracket 35-38 years in the year 2003, and seek to locate their parental 'family market income' decile approximately twenty years earlier. We then compute transition probabilities for the top tail of this

Figure 3: *Intergenerational Earnings Elasticities*

Within and cross-country variations in reported generational earnings elasticities for fathers and sons.



Note: Each vertical bar represents the value of a reported earnings elasticity.

Source: Corak (2004).

distribution of young earners – the top decile, the top 1 per cent and the top 0.1 per cent.

Our strategy obviously does not uncover the family background of all the young high-earning individuals. For example, they may come from immigrant families who were not domiciled in Canada in the early eighties, or their parents may not have filed a tax return in the early eighties. Table 1 contains the first set of results. Each number in that table is rounded to a 0 or 5. Consequently, the zero elements that appear in the top fractile column can be interpreted as ‘small number’ rather than as a strict zero. It is evident that low-value cells, such as those containing ‘0’ or ‘5’, limit the conclusions that can be drawn.

We identified 370,205 individuals in the 2003 LAD sample in the age group 35-38 years, and then attempted to locate each of these individuals in at least one of the years 1982, 1983, 1984. Approximately half were not located (182,865), and some of those located were living as independents rather than in a family context (56,685). This means that the family background was identified for about 35 per cent of the 2003 distribution of young earners. For each year 1982, 1983, 1984 all families in the LAD were ranked by income decile. When the family in which a 2003 earner resided was identified, the decile rank of the family was associated with the young earner. The parental

decile search was started in 1984 and continued back to the earlier years if the individual was not located in the later of the three base years.

There are two well-known challenges in getting reliable estimates of intergenerational income correlations. The first was initially addressed by Solon (1992), who showed that the intergenerational earnings correlation coefficient is biased downward if the annual observed income of the parent differs from the parent's permanent income. The results in Table 1 are based on a single year's parental income. Using multiple years for the parental income is feasible, though it would result in a loss of sample size. Nonetheless, there are two reasons why the bias in our results is likely minimal. First, we use *family* income rather than individual parental income or individual parental earnings. The use of a family measure reduces the year-to-year variation compared to an individual-based measure. In addition, income is a broader measure of household well-being than earnings alone.

The second challenge concerns the point in the lifecycle at which the parents are observed. Graue (2006) shows that this is a key element in reconciling the differing estimates of intergenerational mobility found in US data. To address the possibility that age rather than socio-economic background may be driving the results, we estimated the average decile age of the parents of each of the four quantiles in Table 1 (this was not possible for much of the very top quantile on account of the small numbers involved). Two

Table 1: *Unadjusted Transition Matrix: Parental Income – Offspring Earnings*

<i>Parental Family Market income</i>	<i>Offspring Earnings Quantile</i>				<i>Total</i>
	<i>1–90 (low)</i>	<i>91–99</i>	<i>99.1–99.9</i>	<i>99.91–100</i>	
Decile 1 (low)	3,700	205	15	0	3,920
Decile 2	6,200	435	30	0	6,665
Decile 3	7,395	670	35	5	8,105
Decile 4	8,435	830	65	0	9,330
Decile 5	9,655	1,020	65	10	10,750
Decile 6	10,760	1,265	95	5	12,125
Decile 7	12,595	1,665	135	10	14,405
Decile 8	15,385	2,300	180	15	17,870
Decile 9	18,320	3,405	325	25	22,075
Decile 10 (high)	19,440	5,000	815	130	25,385
Total located in 80s	111,885	16,795	1,760	200	130,630
Not observed in 80s					182,865
Observed in 80s not with parents					56,685
Total 2003					370,205

patterns emerge from this search. First, lower-decile parents tend to be younger than higher-decile parents, as a basic life-cycle model would predict. Despite this, there is very minimal variation in the age of parents within a given decile across earnings quantiles of their children: parents in a given decile whose offspring are high earners are no different age-wise than parents whose offspring are low earners.

The next data challenge relates to the decile size of the located households: Table 1 indicates that many more individuals who were 35-38 years in 2003 were located in the upper deciles of the eighties distribution than in the lower deciles. This is illustrated in the final column. For example, the ninth decile of parents contains 17 per cent of located young earners (22,075/130,630), while the second-from-bottom decile contains just 5 per cent (6,665/130,630).

The lower representation from lower-decile parents means that the offspring of such parents were less likely to be filing a tax return while still domiciled in the parental home. Whether this means they were less likely to work, or less likely to have an income sufficient to warrant filing, or whether they have disproportionately left home (and hence among the 56,685 individuals we have located in this category) cannot be established. Teenagers in Canada tend to apply for a social insurance number when they begin to work. Despite this pattern, we show below that this differential retrieval rate should not affect the main conclusions.

The transitions data in Table 1, unadjusted for differential retrieval rates, indicate that the top 0.1 per cent of earners comes disproportionately from top decile families. Of the 200 whose origins were traced, 130 come from the top decile. Even allowing for the different retrieval rates for the parental deciles, there is a substantial difference in the probability of a top earner coming from the top parental decile as opposed to the ninth parental decile. The adjusted probability is approximately four and a half times higher: if the 25 individuals whose parents are in the ninth decile are weighted by the relative recovery ratio (25,385/22,075) the 25 becomes 29 and therefore the relative probability of a top one thousand earner coming from the top parental decile is 4.5 times the probability of coming from the next decile ($130/29 = 4.5$). Yet the probability of coming from the ninth parental decile is not very different from the probability of coming from any of the fifth to eighth deciles. Sample sizes here are small and caution is obviously appropriate.

The transition probabilities based on rescaled values are presented in Table 2. For illustration: the retrieval rate for decile 10 was 1.72 times the retrieval rate for decile 7 ($25,385/14,405 = 1.72$). Scaling all numbers in Table 1 by their own retrieval rates and calculating the transition probabilities based on these scaled values yields the probabilities in Table 2.

Table 2: *Transition Probabilities Adjusted for Retrieval Differentials*

<i>Parental Family Market Income</i>	<i>Offspring Earnings Quantile</i>			
	<i>1–90 (low)</i>	<i>91–99</i>	<i>99.1–99.9</i>	<i>99.91–100</i>
Decile 1 (low)	0.11	0.05	0.04	0.00
Decile 2	0.11	0.06	0.05	0.00
Decile 3	0.10	0.08	0.04	0.06
Decile 4	0.10	0.08	0.07	0.00
Decile 5	0.10	0.09	0.06	0.10
Decile 6	0.10	0.10	0.08	0.04
Decile 7	0.10	0.11	0.09	0.07
Decile 8	0.10	0.12	0.10	0.09
Decile 9	0.09	0.14	0.15	0.12
Decile 10 (high)	0.09	0.18	0.32	0.53

Our first conclusion is, therefore, that even accounting for different retrieval rates, an earner in the top 0.1 per cent of the distribution has a greater than 50 per cent probability of coming from the top decile of the parental income distribution, but we cannot discern meaningful differences in the probability of coming from immediately lower deciles.

In contrast, for the remainder of the top 1 per cent of the young earners distribution, the relative probability of coming from the top parental decile is much smaller – about twice as high as coming from the ninth decile and three times as high as the probability of coming from the eighth decile. Proceeding down the parental distribution, the probabilities of coming from these lower deciles decline more smoothly than in the case of the top 0.1 per cent. Considering finally the origins of those in the remaining nine centiles within the top decile, the pattern is much more akin to what is observed for the 99.1 – 99.9 fractile than for the 99.91 – 100 fractile: a relatively smooth and declining probability function.

In sum, the pattern defining the socio-economic origins of the extreme top of the earnings distribution is different from the patterns defining the remaining parts of the top decile of young earners. For those in the top one thousandth of the distribution, the probability of coming from the top decile of the parental distribution is in excess of 50 per cent, and the probabilities of coming from other deciles do not decay smoothly as they do with the other parts of this top earnings decile.

The sample size was not sufficiently large to permit a breakdown of these transition probabilities between men and women. Only about 10 per cent of the top 0.1 per cent are women. From a total of 200 individuals retrieved for

this fractile the percentage formed by women is much too small given that individuals must be allocated across all parental deciles.

Finally, there is the possibility that the difference in retrieval rates across parental deciles may bias the results. In Table 2 the retrieved data were scaled assuming that the 'missing' individuals come from the parental distribution in the same proportion as the observed individuals. However, it is more likely that the 'undiscovered' individuals come disproportionately from the lower parts of the parental decile. Those individuals not included in the 130,630 belong to three main groups: children of parents who migrated to Canada after 1984; children observed to be living outside of the parental household (numbering 56,685) and a large remaining number, most of whom would have been living in the parental household without a social insurance number.

It is difficult to say anything definite about the likely placement of those of immigrant origin. However, for the second group – those individuals living independently in their late teens, it is less likely that they went to university or obtained third level education than those who continued to live in the parental household, and therefore it is less probable that they would show up in the top part of the earnings distribution of 35-38 year olds in 2003. The largest missing group is likely those with no social insurance number and still living at home in the early eighties. Since most teens in Canada obtain a SIN when they first take a temporary job, our 'retrieved' individuals are likely those who have some employment income while a teen and living at home, while the 'unretrieved' individuals have not a SIN and hence unlikely to have income as a teen. For our results to be overstated we would have to believe that teens who have no income are more likely to be successful later in life than teens who do have some income. Such an argument is difficult to support.

3.2 *Top Fractile Representation of Women*

A body of literature now exists that focuses upon wage disparities between men and women at different parts of the distribution. A 'glass ceiling' is the term that is frequently used to describe a situation when the conditional distribution of earnings for one group – in this case women – differs from that of another – men. Albrecht *et al.* (2003) is an example. They find this kind of disparity characterises the distributions in Sweden.

While our prime interest is in the extreme top of the distribution, findings for four components of the distribution are presented in Table 3: the top 0.1 per cent, the next 0.9 per cent, the following 9 per cent and then the 'bottom' 90 per cent. This enables us to compare the representation of women with men at different points at the top of the distribution. Table 3 contains the proportion of the quantiles composed of women for a series of years between 1983 and 2003. We have computed the statistics for every year, but since the

Table 3: Earnings Distributions Characteristics by Gender (contd)

Year	Quantile	Count	Quantile Sum	Female / Male Numbers	Mean Earnings	Female / Male Earnings	Catch-up Index (alpha)		
							0.2	0.5	0.8
1989	0-90	F 1,354,570			12,300				
	0-90	M 1,134,265	2488835	1.19	20,600	0.60			
	90.1-99	F 35,285			55,100				
	90.1-99	M 213,600	248885	0.17	57,300	0.96	0.77	0.60	0.32
	99.1-99.9	F 2,270			128,600				
	99.1-99.9	M 22,620	24890	0.10	133,600	0.96	0.84	0.69	0.39
	99.91-100	F 185			509,900				
	99.91-100	M 2,580	2765	0.07	574,900	0.89	0.88	0.75	0.46
	99.91-100	M 2,580							
1991	0-90	F 1,394,390			13,100				
	0-90	M 1,182,740	2577130	1.18	19,900	0.66			
	90.1-99	F 46,485			59,000				
	90.1-99	M 211,230	257715	0.22	61,800	0.95	0.70	0.54	0.29
	99.1-99.9	F 2,545			130,600				
	99.1-99.9	M 23,230	25775	0.11	136,700	0.96	0.83	0.68	0.38
	99.91-100	F 170	413100						
	99.91-100	M 2,695	2865	0.06	491,900	0.84	0.89	0.77	0.50
	99.91-100	M 2,695							
1993	0-90	F 1,458,190			13,200				
	0-90	M 1,232,115	2690305	1.18	19,500	0.68			
	90.1-99	F 55,645			61,500				
	90.1-99	M 213,385	269030	0.26	64,200	0.96	0.66	0.50	0.26
	99.1-99.9	F 2,945			137,400				
	99.1-99.9	M 23,955	26900	0.12	143,800	0.96	0.81	0.66	0.37
	99.91-100	F 195			475,100				
	99.91-100	M 2,795	2990	0.07	560,400	0.85	0.89	0.76	0.49
	99.91-100	M 2,795							

Table 3: Earnings Distributions Characteristics by Gender (contd.)

Year	Quantile	Count	Quantile Sum	Female/Male Numbers		Mean Earnings	Female/Male Earnings	Catch-up Index (alpha)			
				Female	Male			0.2	0.5	0.8	
1995	0-90	1,484,595				13,800					
	0-90	1,259,700	2744295	1.18		20,300	0.68				
	90.1-99	55,475				64,100					
	90.1-99	218,970	2744445	0.25		67,300	0.95	0.67	0.51	0.27	
	99.1-99.9	3,340				150,200					
	99.1-99.9	24,110	27450	0.14		158,500	0.95	0.80	0.64	0.35	
	99.91-100	215				539,100					
1997	99.91-100	2835	3050	0.08		606,800	0.89	0.88	0.74	0.46	
	0-90	1,495,185				14,800					
	0-90	1,272,800	2767985	1.17		21,600	0.69				
	90.1-99	56,380				68,500					
	90.1-99	220,425	276805	0.26		71,800	0.95	0.67	0.51	0.27	
	99.1-99.9	3,685				177,300					
	99.1-99.9	23,995	27680	0.15		185,900	0.95	0.78	0.62	0.34	
1999	99.91-100	220				686,600					
	99.91-100	2,855	3075	0.08		827,500	0.83	0.88	0.75	0.48	
	0-90	1,513,720				16,400					
	0-90	1,284,415	2798135	1.18		23,500	0.70				
	90.1-99	60,325				74,600					
	90.1-99	219,495	279820	0.27		77,800	0.96	0.65	0.49	0.25	
	99.1-99.9	3,975				199,600					
1999	99.1-99.9	24,005	27980	0.17		210,300	0.95	0.77	0.60	0.33	
	99.91-100	275				825,000					
	99.91-100	2,835	3110	0.10		989,200	0.83	0.85	0.72	0.46	

Table 3: *Earnings Distributions Characteristics by Gender (contd.)*

Year	Quantile	Count	Quantile Sum	Female/ Male Numbers	Mean Earnings	Female/ Male Earnings	Catch-up Index (α)		
							0.2	0.5	0.8
2001	0-90	F	1,536,650		18,500				
	0-90	M	1,322,725	1.16	25,900	0.71			
	90.1-99	F	65,670		82,300				
	90.1-99	M	220,270	0.30	86,000	0.96	0.62	0.47	0.24
	99.1-99.9	F	4,445		231,900				
	99.1-99.9	M	24,150	0.18	243,300	0.95	0.74	0.58	0.31
	99.91-100	F	285		1,058,600				
	99.91-100	M	2,895	0.10	1,237,100	0.86	0.85	0.71	0.44
	2003	0-90	F	1,549,380		19,600			
0-90	M	1,335,560	1.16	27,000	0.73				
90.1-99	F	72,110		86,700					
90.1-99	M	216,410	0.33	90,800	0.95	0.59	0.44	0.23	
99.1-99.9	F	4,750		237,100					
99.1-99.9	M	24,100	0.20	249,200	0.95	0.73	0.57	0.31	
99.91-100	F	320		1,012,900					
99.91-100	M	2,885	0.11	1,173,500	0.86	0.83	0.69	0.43	

numbers are not highly cyclical, just a subset of the years is presented. The smallest sample in any year in this table is of the order of two and a half million individuals.

The 'female/male numbers' column contains the percentage of each specified quantile that is composed of females. This ratio attains a value of 0.11 for the top 0.1 per cent of the distribution in the most recent year of data availability – 2003 – starting from 0.05 in 1983. Its ascent has been gradual, with a stumble that coincided with the early nineties recession. The growth of this statistic in the remaining part of this top centile has been more rapid: it has increased from just over 0.06 to 0.20. Furthermore, the share appears not to have been impacted by the nineties recession. The ratio in the remainder of the top decile (excluding the top 1 per cent) has increased from 0.14 to 0.33. (Of course, the *share* of females in these deciles is smaller than the foregoing fractions, being the ratio of the number of females to the total, rather than to the number of males).

Evidently these percentages are small, and rates of share growth are low. At the same time it is not the case that progress is only outside of the top centile. There may be a glass ceiling, but it is receding slowly throughout the whole upper decile.

These participation rates tell us nothing about the distribution of male and female *earnings* within the various quantiles. In particular, the very top of the distribution is likely characterised by a heavily skewed, perhaps Pareto, tail and it may be optimistic to believe that women and men are distributed in the same manner throughout this tail. We have, therefore, computed the average income level of men and women in each of the quantiles. These are given in the next column headed 'mean earnings' and the following column 'female/male earnings' defines the relative mean earnings values. Interestingly, the shortfall of female earnings is small and there is no strong pattern over time for this top 0.1 per cent of earners. The same pattern holds true for the remainder of the top centile and the remainder of the top decile.

Since the averages for women are generally below the averages for men, the lower participation rate for women than men in essence understates what might be termed the 'income representation' of women. In the poverty literature, measures exist that capture both the rate of poverty and the income shortfall of the poor – for example Sen (1976), Foster, Greer and Thorbecke (1984) and Shorrocks (1995). But, to our knowledge, there are no indices that combine these dimensions at the top of the distribution. The Sen and Shorrocks indices are essentially the product of a headcount index and an average income shortfall index, weighted linearly. They are not readily adaptable to the top tail however, because the 'excess income' beyond a threshold defined by a fractile boundary, relative to the threshold, is not

constrained to be less than one.

A measurement solution suggests itself, which we term a ‘catch-up’ index. This measures both the shortfall in representation of one group relative to another in the high echelons of the distribution, and also the degree to which the incomes of the less favoured group fall short of the incomes of the favoured group. Define the proportion of females in the top group by H_f and the proportion of men by H_m – analogous to the headcount measure used in poverty analysis. Let the mean income of each group be μ_f and μ_m . Then as long as $\mu_m > \mu_f$ and $H_m > H_f$ the index

$$I = 1 - \frac{\mu_f^\alpha H_f^{(1-\alpha)}}{\mu_m^\alpha H_m^{(1-\alpha)}}$$

is bounded between zero and one for alpha lying on the unit interval, where alpha represents one’s aversion to representation relative to income shortfall.

The index values are given in the final columns of Table 3 for different values of the sensitivity parameter. There is a general decline in these over time reflecting the increased participation of women in the top fractiles. The index that places a larger weight on earnings than participation has a value closer to zero, indicating more equality between men and women, since the low participation of women in these fractiles is heavily discounted.

3.3 Spousal Incomes

A frequently proposed reason for why women are under-represented in the very top fractiles is that they have less social and familial support from their partners. While many men in the top fractiles have female partners who are full-time supporters and family managers, this is much less true of women. Since the LAD contains separate information on the earnings for each individual in the household, it is an ideal database for investigating this hypothesis, and to establish the degree to which patterns have changed since 1982.

While tax payers file on an individual basis in Canada, the LAD also matches individuals with their partners using a variety of matching criteria. The samples upon which Table 4 are based include individuals in the same age group as above, and the mean spousal earnings figures are for those spouses who have some minimal connection with the labour market – in this case we set a minimum earnings level of \$1,000 for a spouse to be defined to have earnings.

The final two columns in Table 4 indicate that spousal earnings patterns are very asymmetric in the top fractiles. Women in the top 0.1 per cent have husbands with positive earnings in about 90 per cent of cases, and this

Table 4: *Spousal Earnings Patterns*

		<i>Quantile</i>	<i>Quantile Count</i>	<i>Quantile Sum</i>	<i>Mean Earnings</i>	<i>Spousal Mean Earnings</i>	<i>Per Cent Positive</i>
1983	F	0-90	789,500		7,300	25,500	0.88
	M	0-90	389,420	1,178,920	16,600	12,400	0.36
	F	90.1-99	14,305		38,300	39,300	0.87
	M	90.1-99	92,500	106805	40,200	15,200	0.19
	F	99.1-99.9	740		77,800	105,700	0.89
	M	99.1-99.9	9,530	10,270	79,100	19,100	0.06
	F	99.91-100	75		220,200	180,900	0.87
	M	99.91-100	12,05	1,280	234,300	32,500	0.28
1985	F	0-90	776,520		8,600	28,600	0.88
	M	0-90	409,275	1,185,795	18,700	13,600	0.45
	F	90.1-99	14,965		42,800	44,800	0.86
	M	90.1-99	99,010	1,139,75	44,600	16,700	0.30
	F	99.1-99.9	915		87,800	108,600	0.83
	M	99.1-99.9	10,410	11,325	89,100	21,700	0.23
	F	99.91-100	80		291,100	410,100	0.81
	M	99.91-100	1,300	1,380	277,800	38,400	0.36
1987	F	0-90	799,920		10,100	31,300	0.88
	M	0-90	448,145	1,248,065	20,800	15,000	0.54
	F	90.1-99	18,490		47,300	51,600	0.88
	M	90.1-99	111,135	129,625	49,200	18,900	0.45
	F	99.1-99.9	1,265		99,600	126,700	0.86
	M	99.1-99.9	11,685	12,950	101,800	25,900	0.39
	F	99.91-100	105		254,000	232,000	0.81
	M	99.91-100	1,410	1,515	339,500	46,600	0.46
1989	F	0-90	842,060		12,200	35,400	0.87
	M	0-90	516,380	1,358,440	23,900	17,000	0.60
	F	90.1-99	21,330		55,300	65,300	0.90
	M	90.1-99	128,640	149,970	57,200	22,100	0.56
	F	99.1-99.9	1,515		129,000	153,200	0.89
	M	99.1-99.9	13,595	15,110	133,900	32,100	0.51
	F	99.91-100	145		433,900	815,700	0.93
	M	99.91-100	1,590	1,735	602,300	68,700	0.52
1991	F	0-90	818,335		13,500	36,700	0.83
	M	0-90	520,745	1,339,080	23,800	18,900	0.58
	F	90.1-99	27,840		59,200	64,200	0.88
	M	90.1-99	129,390	157,230	61,800	25,500	0.60
	F	99.1-99.9	1,670		131,600	139,700	0.89
	M	99.1-99.9	14,035	15,705	136,900	34,900	0.52
	F	99.91-100	130		422,300	472,700	0.92
	M	99.91-100	1,620	1,750	491,500	61,600	0.50
1993	F	0-90	84,5825		14,200	37,400	0.80
	M	0-90	54,7790	1,393,615	24,000	20,200	0.55
	F	90.1-99	34040		61700	62900	0.87
	M	90.1-99	133635	167675	64200	27600	0.62

Table 4: *Spousal Earnings Patterns (contd.)*

		<i>Quantile</i>	<i>Quantile</i>	<i>Quantile</i>	<i>Mean</i>	<i>Spousal</i>	<i>Per Cent</i>
		<i>Quantile</i>	<i>Count</i>	<i>Sum</i>	<i>Earnings</i>	<i>Mean</i>	<i>Positive</i>
						<i>Earnings</i>	
1993	F	99.1-99.9	1,990		137,600	147,600	0.90
contd.	M	99.1-99.9	14,850	16,840	143,700	38,000	0.55
	F	99.91-100	135		490,700	430,200	0.89
	M	99.91-100	1,605	1,740	546,300	66,600	0.44
1995	F	0-90	852,805		15,000	39,300	0.80
	M	0-90	555,805	1,408,610	25,000	21,000	0.55
	F	90.1-99	34,380		64,400	67,900	0.87
	M	90.1-99	137,615	171,995	67,300	28,700	0.64
	F	99.1-99.9	2,245		150,800	155,700	0.89
	M	99.1-99.9	15,245	17,490	158,700	41,500	0.59
	F	99.91-100	155		548,800	502,600	0.90
	M	99.91-100	1,655	1,810	600,800	75,200	0.44
1997	F	0-90	870,380		16,000	41,900	0.80
	M	0-90	577,000	1,447,380	26,400	22,100	0.58
	F	90.1-99	35,635		68,800	77,400	0.88
	M	90.1-99	139,780	175,415	71,900	30,600	0.65
	F	99.1-99.9	2,515		177,600	157,100	0.89
	M	99.1-99.9	15,360	17,875	185,600	46,800	0.61
	F	99.91-100	160		683,500	431,000	0.88
	M	99.91-100	1,710	1,870	800,800	78,600	0.48
1999	F	0-90	878,470		17,700	45,000	0.81
	M	0-90	601,800	1,480,270	28,400	23,900	0.62
	F	90.1-99	38,545		74,900	84,100	0.89
	M	90.1-99	141,560	180,105	77,700	33,700	0.69
	F	99.1-99.9	2,715		199,700	174,900	0.89
	M	99.1-99.9	15,450	18,165	209,300	52,700	0.62
	F	99.91-100	205		791,800	535,700	0.90
	M	99.91-100	1,640	1,845	971,300	99,000	0.46
2001	F	0-90	903,460		19,700	48,800	0.83
	M	0-90	638,100	1,541,560	30,800	26,000	0.67
	F	90.1-99	42,545		82,700	93,000	0.90
	M	90.1-99	144,105	186,650	86,000	37,300	0.72
	F	99.1-99.9	3,040		233,900	210,000	0.90
	M	99.1-99.9	15,525	18,565	242,000	59,600	0.62
	F	99.91-100	195		1,008,700	698,800	0.90
	M	99.91-100	1,690	1,885	1,180,000	102,200	0.46
2003	F	0-90	903,980		21,000	50,900	0.83
	M	0-90	649,000	1,552,980	32,300	27,900	0.68
	F	90.1-99	46,520		87,100	90,600	0.89
	M	90.1-99	143,655	190,175	90,700	39,800	0.73
	F	99.1-99.9	3,315		236,800	188,900	0.90
	M	99.1-99.9	15,525	18,840	248,100	62,400	0.62
	F	99.91-100	240		1,028,300	730,100	0.90
	M	99.91-100	1,660	1,900	1,132,900	119,000	0.43

number is relatively constant throughout the sample period. In contrast, men have wives with positive earnings in about 50 per cent of cases in the modern era, and a somewhat lower percentage in the early eighties.

These patterns are broadly similar to those in the remainder of the top percentile, though in the latter case a higher number of female spouses have positive earnings – about 60 per cent rather than 50 per cent. But the figure for male spouses in this fractile is similar to the male figure recorded in the top 0.1 per cent. The first part of the conventional wisdom is therefore borne out by these data.

The actual earnings data indicate that the asymmetry between men and women is much more severe at the top compared with the asymmetry in the participation rates. Women in the top 0.1 per cent of the distribution have partners who earn about 70 per cent of the female earnings in the most recent period, and about 100 per cent of female earnings in the eighties. For example, in 2003 the top 0.1 per cent of women earned just in excess of one million dollars while their partners earned about \$730,000. In contrast, men in the top 0.1 per cent have spouses who, when in the labour market, earn about one-tenth of the male figure. For example, in 2003 the respective figures were \$1.13m and \$119,000.

The asymmetry is slightly less severe for the remaining part of the top centile – men have spouses who earn about one-fifth of male earnings when working. What these earnings figures indicate about hours of work, this data base does not allow us to ascertain, even if it is recognised that labour supply decisions within the household are made jointly. Nor does our data base have any information on educational attainment, which is one marker of the degree of ‘assortative mating’ in the population.

3.4 *The Post ‘Tech Bust’ Era*

As a postscript to this analysis it is worth noting that the earnings shares of the top fractiles have not declined in the post-2000 era. If the growing share of the top one thousandth of the distribution is being driven by the earnings patterns of individuals working in corporations, the tech-bust of 2000 has not had a perceptible impact on this growth. In addition, like Saez and Veall, we have investigated the ‘permanent’ incomes of those at the top by taking three-year moving averages of earnings. Gottschalk and Moffitt (1994) proposed that some of the increase in US inequality in the modern era is attributable to a greater variance in transitory incomes. However, such an averaging has virtually no impact on the share growth at the very top for the whole period that we have examined, *nor does it have any impact in the post-2000 era.*

IV CONCLUSIONS

This paper investigates some of the socio-economic characteristics of the new class of Canadian ‘super earners’ that has emerged in the last two decades. Virtually all other English-speaking economies, including Ireland, have also witnessed the emergence of a top group whose incomes have increased dramatically. We have been fortunate to have at our disposal a data base of about eighty million tax filers that has yielded large reliable samples at the very extremity of the earnings distribution. Our principal findings are: first, the top one thousandth of the distribution comes disproportionately from the top decile of the parental distribution, whereas the remaining component of the top centile is considerably more likely to have its origins in the parental deciles below the top. Second, women have a low, but increasing, rate of participation in the very top tail; they now form about 10 per cent of the group, as opposed to 5 per cent in the early eighties. In contrast, the earnings of women in this group are not so different from male earnings. Third, there is a very severe asymmetry in the earnings patterns of male and female spouses. About half of top male earners have working partners, whereas about 90 per cent of top female earners have working partners. Furthermore, the earnings of ‘secondary males’ are between five and ten times the earnings of ‘secondary females’ in the top 0.1 per cent of the distribution. Fourth, the concentration of earnings at the top tail is not attributable to any growth in the transitory component of earnings, nor was there any break in this pattern with the ‘tech bust’ of 2000.

All data bases have limitations, and the LAD is no exception. While it is rich in the income dimension it has no information on human capital or occupation. This clearly limits the inferences that can be drawn from it. In addition, it is clear that we cannot say if these patterns that we have established for Canada are mirrored in the US, the UK, New Zealand, Australia or Ireland. But they represent a step in filling some of the holes in our knowledge of the characteristics of those individuals and households who inhabit the very top tail of the earnings distribution in Canada.

REFERENCES

- ALBRECHT, J., A. BJORKLAND and S. VROMAN, 2003. “Is There a Glass Ceiling in Sweden?” *Journal of Labor Economics*, Vol. 21, No. 1, pp. 145-177.
- ALESINA, A., R. BAQIR and W. EASTERLY, 1999. “Public Goods and Ethnic Divisions”, *Quarterly Journal of Economics*, Vol. 114, pp. 1243-1284.
- ALVAREDO, F., E. SAEZ, 2005. “Income and Wealth Concentration in Spain in a Historical and Fiscal Perspective”, Working Paper, Department of Economics,

- University of California at Berkeley, September.
- ATKINSON, A., 2005. "Top Incomes in the United Kingdom over the Twentieth Century", *Journal of the Royal Statistical Society*, Series A, Part 2, pp. 325-343.
- ATKINSON, A., F. BOURGUIGNON and C. MORRISON, 1992. *Empirical Studies of Earnings Mobility*, Switzerland: Harwood Academic Publishers.
- ATKINSON, A. and A. LEIGH, 2004. "Understanding the Distribution of Top Incomes in Anglo-Saxon Countries over the Twentieth Century", Department of Economics, *Australian National University*, November.
- ATKINSON, A. and W. SALVERDA, 2005. "Top Incomes in the Netherlands and the UK in the Twentieth Century", *Journal of the European Economic Association*, Vol. 3, pp. 883-913.
- BANERJEE, A. and T. PIKETTY, 2003. "Top Indian Incomes, 1922-2000", CEPR Discussion Paper, # 4632, 2003.
- BUCHINSKY, M. and J. HUNT, 1999. "Wage Mobility in the United States", *Review of Economics and Statistics*, Vol. 81, No. 3, pp. 351-368.
- CORAK, M., 2004. "Do Poor Children Become Poor Adults? Lessons for Public Policy from a Cross Country Comparison of Generational Earnings Mobility", Discussion Paper, UNICEF Innocenti Research Centre, Florence, Italy, September .
- CORAK, M. and A. HEISZ, 1999. "The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax Data", *Journal of Human Resources*, Vol. 34, pp. 504-533.
- DELL, F., T. PIKETTY and E. SAEZ, 2005. "Income and Wealth Concentration in Switzerland over the 20th Century", CEPR Discussion Paper 5090, May.
- DIAMOND, J., 2005. "Collapse. How Societies Choose to Fail or Succeed", New York: Viking Press.
- DUCLOS, J.Y., J. ESTEBAN and D. RAY, 2004. "Polarisation: Concepts, Measurement, Estimation", *Econometrica*, Vol. 72, pp. 1737-1772.
- ECKSTEIN, Z. and E. NAGYPAL, 2002. "The Evolution of U.S. Earnings Inequality: 1961-2002", *Federal Reserve Bank of Minneapolis, Quarterly Review*, December.
- FAJNZYLBER, P., D. LEDERMAN and N. LOAYZA, 2002. "Inequality and Crime", *Journal of Law and Economics*, Vol 45, No. 1.
- FELDSTEIN, M., 1995. "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act", *Journal of Political Economy*, Vol. 103, No. 3, pp. 551-572.
- FOSTER, J., J. GREER and E. THORBECKE, 1984. "A Class of Decomposable Poverty Measures", *Econometrica*, Vol. 52, No. 3, pp. 761-766.
- GLAESER, E., 2005. "Inequality", Harvard Institute of Economic Research, Discussion Paper Number 2078, July.
- GOTTSCHALK, P. and R. MOFFITT, 1994. "The Growth of Earnings Instability in the US Labor Market", *Brookings Papers on Economic Activity*, pp. 217-272.
- GRAWE, N., forthcoming, 2006. "The Extent of Lifecycle Bias in Estimates of Intergenerational Earnings Persistence", *Labour Economics*.
- KENNICKELL, A., 2001. "An Examination of Changes in the Distribution of Wealth from 1989 to 1998: Evidence from the Survey of Consumer Finances", Working Paper, Washington: Federal Board.
- MORIGUCHI, C. and E. SAEZ, 2005. "The Evolution of Income Concentration in Japan, 1885-2002: Evidence from Tax Statistics", Working Paper, Department of Economics, University of California at Berkeley.

- NOLAN, B., 2006. "Long-Term Trends in Top Income Shares in Ireland", in A. B. Atkinson and T. Piketty (eds.), *Top Incomes over the Twentieth Century*, Oxford: Oxford University Press.
- PIKETTY, T., 2003. "Income Inequality in France, 1901-1998", *Journal of Political Economy*, Vol. 111, No. 5, pp. 1004-1042.
- PIKETTY, T. and E. SAEZ, 2003. "Income Inequality in the United States, 1913-1998", *Quarterly Journal of Economics*, Vol. 118, No. 1, pp. 1-39.
- SALA-I-MARTIN, X., 2002. "The World Distribution of Income", Columbia University, Economics Department Working Paper, April.
- SAEZ, E. and M. VEALL, 2005. "The Evolution of High Incomes in Northern America: Lessons from Canadian Evidence", *American Economic Review*, Vol. 95, No. 3, pp. 831-849.
- SEN, A., 1976. "Poverty an Ordinal Approach to Measurement", *Econometrica*, Vol. 44, pp. 219-231.
- SHORROCKS, A., 1995. "Revisiting the Sen Poverty Index", *Econometrica*, Vol. 63, pp. 1225-1230.
- SOLON, G., 1992. "Intergenerational Income Mobility in the United States", *American Economic Review*, Vol. 82, pp. 393-408.
- SOLON, G., 1989. "Biases in the Estimation of Intergenerational Earnings Correlations", *Review of Economics and Statistics*, Vol. 71, pp. 172-174.