Earnings Biases in the United Kingdom Regional Accounts: Some Economic Policy and Research Implications

Gavin Cameron and John Muellbauer
Nuffield College, Oxford
14 December 1999

Abstract:

Between the late 1970s and the late 1980s, the United Kingdom Regional Accounts data suggest a much smaller rise in the South East earnings premium and consequently a much smaller increase in the regional dispersion of earnings than do the other sources of data on regional earnings. The paper discusses a number of possible explanations for this discrepancy and concludes that it was probably due to problems at the Inland Revenue in allocating tax records across the regions. The historical unreliability of the Regional Accounts has possible implications for economic research on both regional consumption and regional convergence and may have caused the poorest regions to miss out on EU Structural Funds.

* This research was funded by ESRC grant number R00023 7500, ‘Modelling Non-Stationarity in Economic Time Series’. We are particularly grateful for patient help and advice on data issues to Tim Holt, Derek Bird, Jude Hillary and Dev Virdee, as well as Dorothy Anderson, Guy Manley, James Partington, Jane Roberts, and David Wilkinson, and to David Heald, Iain McLean, Jorg Mathias, Mike Phelps and Mark Stephens for helpful discussions. We take full responsibility for any errors.
This paper argues that United Kingdom Regional Accounts data on income from employment and hence personal income and gross domestic product per head for the last 20 years give a misleading picture of the development of regional differentials. These data are published in Economic Trends and Regional Trends and are widely used. They have informed public debates about such topics as Scottish and Welsh devolution. They are an important input for policy towards the English regions and for Wales, Scotland and Northern Ireland and are used to define the income criteria by which a region or area qualifies for assistance from the European Union.\(^1\) The regional income data have also been widely used by economists to draw conclusions on such issues as whether there are tendencies for regional incomes to converge and the relative roles of, for example, income, housing wealth and unemployment in determining regional consumption. In turn, research findings on such issues can influence policy as well as the general climate of economic ideas.

The paper is structured as follows. Regional Accounts data on income from employment and on wages and salaries are, in aggregate, consistent with the National Accounts. Section 1 therefore compares National Accounts data on earnings with alternative sources such as the Average Earnings Index, the New Earnings Survey and the Family Expenditure Survey. This throws light on possible differences caused by data sources and methods of construction. Section 2 focuses on the differential for various measures of average earnings between the South East and Great Britain as a whole.\(^2\) The Regional Accounts and Family Expenditure Survey data are based upon the place of residence, while the New Earnings Survey data are based upon the place of employment. The Regional Accounts data suggest a much smaller rise in the South East premium between 1979 and 1987 than do the other sources on regional earnings. Why the

---

\(^1\) Under the 1989-1993 and 1994-1999 European Community Structural Funds regulations, regions at NUTS level II whose per capita GDP, in a three year qualifying period which is typically four years before the funding period, is less than 75 per cent of the Community average receive Objective I status and funding.

\(^2\) Since the full New Earnings Survey results for Northern Ireland are not published, this paper focusses on comparisons between the South East and Great Britain. In 1994, Northern Ireland accounted for 2.4 per cent of the United Kingdom labour force.
regional accounts data differ from other sources is a puzzle. However, Section 3 rejects the hypothesis that differences between place of residence and place of employment based data can explain much of the anomalous behaviour of the Regional Accounts income data. It also rejects the hypothesis that shifts in the structure of employment towards part-time work are a significant part of the explanation. Instead, the analysis suggests that problems in the methods used by the Inland Revenue and the Office of National Statistics to apportion national measures of income from employment to the regions are the most important part of the explanation for the anomalous behaviour of the Regional Accounts income data. Section 4 summarizes and discusses some of the implications of our findings.

1. A Comparison of Earnings Data for Great Britain

In the National Accounts, estimates of income from employment and of wages and salaries come from a one percent sample of tax and social security records combined with estimated earnings for those below the relevant tax and social security contributions floors and some adjustments for other suspected reporting biases which are used to balance the national accounts. Income from employment is defined as wages and salaries plus employers’ National Insurance and occupational pensions contributions. It is then possible to obtain an estimate of average earnings per job by dividing total income from employment or wages and salaries by the number of employees in employment. The latter is a count of full-time and part-time jobs.

Perhaps the most systematic alternative source of data on average earnings is the New Earnings survey. Average earnings can be constructed from the New Earning Survey (NES) by weighting average earnings of workers split by gender, manual/non-manual and full-time/part-time characteristics by employment shares of these different types of workers.
The NES is also largely based on a one percent sample of national insurance records obtained from PAYE ('pay as you earn') records ending in a specified pair of digits.\textsuperscript{3,4} The use of PAYE records tends to undercount those workers with earnings below the income tax threshold, £77.79 a week in 1997. For example, in 1997, 8.9% of NES employees earned below this amount. This compares with 16.7% of employees in the Labour Force Survey when multiple jobs are taken into account. These workers are mainly part-time women, a small number of young people and others with low weekly earnings. For example, in 1994, the sample of full-time workers in the New Earnings Survey was just over 0.80% compared to 1% expected from a 100% response rate. For part-time workers, only a 0.60% sample was achieved.

The NES earnings data relate to April of each year and exclude employers’ National Insurance and occupational pensions contributions. We have national (and regional) weekly earnings per employment of workers whose pay was not affected by absence for full-time non-manual and manual males and females, for part-time females and part-time males (the latter only for Great Britain).

To obtain average earnings per employment, we weight these by their respective fractions in the total number of jobs, using Census of Employment estimates for full-time and part-time males and females and NES ratios of numbers of non-manuals relative to manuals to split full-time males and females into these categories (since Census of Employment estimates for non-manuals and manuals do not exist). Because of the earnings cut-off, lower paid part-time workers are under-represented in the NES and average part-time earnings are therefore biased.

\textsuperscript{3} Before 1975, the New Earnings Survey was based on a sample of National Insurance Cards which would have had a slightly wider coverage than PAYE records.

\textsuperscript{4} The same digits have been used since 1975 so that there is a large longitudinal element in the sample. From 1983 to 1994 around 75% of the sample came from lists of national insurance records including the identity of the employers and around 25% from lists supplied directly by large organizations of employees with the appropriate national insurance numbers. The latter component of the sample is likely to be more up-to-date since new hires show up in tax office records with a lag and some temporary hires do not show up at all. The use of large organizations in this way seems to have begun in 1981, partly to save costs. The share was unspecified in 1981, 20% in 1982 and around 25% from 1983 to 1994. From 1995, the share has been reduced again to around 12% and in 1997 was down to 10%. It is possible these shifts may have induced small distortions, e.g., in average earnings of part timers.
By using Census of Employment data on the number of part-time women, we correct the employment weight but cannot correct the average earnings figures for part-time women. This procedure should therefore give a slightly upward biased estimate of average earnings. To obtain full year estimates of average earnings, we take the (geometric) average of current April's earning with a weight of 0.75 and of the following April's earnings with a weight of 0.25.

We can now compare these figures for average earnings with the National Accounts income from employment divided by the number of jobs (i.e., employees in employment). Figure 1a shows the log ratio of the National Accounts (NA) estimate to the NES estimate for 1972-1995. Despite the likely upward bias in NES figures, the log-ratio exceeds 0.15 and rises over time. The main reasons for the discrepancy at the national level are likely to be the following:

i. The NA estimate of income from employment includes employers' contributions to National Insurance and occupational pensions, while the NES does not. These are excluded in the National Accounts wages and salaries figures, and dividing these by the number of jobs gives Figure 1b. The above differences between the NA and other sources fall considerably and are fairly stable over the long-run.

ii. The NA estimate covers a full year's data, in particular bonuses and commissions which tend to be paid towards the end of the calendar year and are therefore under-represented in the NES earnings figures. It is believed that these represent a rising fraction of annual earnings.

iii. The NA estimate includes estimates of under-reported earnings (including 'black-economy' adjustments) which help reconcile the data from different sources for the National Accounts. It is widely believed, for example, that earnings of building workers and hotel restaurants workers tend to be under-estimated in official sources.

iv. There could also be a small cyclical discrepancy: The NES data refer to earnings of workers not affected by absence. It seems likely that such earnings will be more stable over the business cycle than average earnings for all workers. Thus, in down-turns, one might expect the NES data to overstate average earnings of all workers. The ratio of
NA:NES wage bills would then fall in business down-turns. However, Figure 1a shows no clear evidence of such a cyclical discrepancy.\footnote{Indeed, the data for 1975, a year of recession, show that National Accounts income data were high relative to all three of the alternative sources.}

If the NES is the most systematic alternative source of date on average earnings, the Average Earnings Index (AEI) is the most obvious. However, coverage of the Average Earnings Index in the 1970s was limited, with the service sector being under-represented. Since then the sectoral coverage has improved. As the temporary suspension of the AEI in 1998 revealed, however, the sampling frame had become somewhat out-of-date in recent years, with new companies being under-represented in the index.\footnote{The King-Turnbull report from HM Treasury and the Chambers report from the Office of National Statistics, published in 1999 and available from the respective web sites, give details. See also Chambers, Weale and Youll (2000).} The AEI data are also shown in Figures 1a and 1b, and show little trend relative to the National Accounts data.

Another systematic source of data on earnings is the Family Expenditure Survey (FES). This is based on a sample of around 7,000 households per annum, reflecting a response rate of around 70%. One advantage of the FES is that it enables average earnings per household to be calculated. Visual examination of the annual FES earnings data show a somewhat volatile series. By taking a three year moving average of the FES data deflated by the consumer price index (using the consumer expenditure deflator), we can find a less volatile estimate of the ratio of National Accounts/ FES average real earnings per household. Figure 1a plots the log-ratio and once again Figure 1b adjusts the NA data by excluding employers' contributions. Despite taking the 3 year moving average, the FES data appear to be more volatile than the other three sources (see Foster, 1998). However, there is little sign of trend divergence between the NA per household earnings figures and the FES figures.\footnote{This conclusion is, however, subject to one qualification. In 1982, there was a shift in the FES in the definition of who was counted as being employed. From 1982, this would have led to an undercounting of earnings by part-time workers working only a few hours per week. This may help explain the downward blip in FES earnings per household in 1982 compared with earlier years.}
2. A Comparison of Regional Earnings

In the United Kingdom, national estimates of income from employment, and since 1984, of wages and salaries are distributed over the regions using data from a one percent sample of national insurance records for a complete tax year (April to March) adjusted to a calendar year basis. These tax records are on a place of residence basis so that the regional wage and salary bill is also on a residence basis.

Let us begin our sequence of comparisons by comparing the log ratio of earnings per household in the South East to those in Great Britain from the Regional Accounts with the three year moving average of those from the FES. Figure 2 shows the results. According to the Regional Accounts, the South East earnings premium shows only around a 2% rise between 1979 and 1987 while the FES data show a rise in the South East premium of around 11%. The smoothness of the FES data in Figure 2 compared with Figure 1a and Figure 1b suggests that the FES may give more reliable estimates of regional differentials for populous regions than of average earnings nationally. Studies of the representativeness of the FES suggest that high earners are under-represented, especially in the South East (see Foster, 1998). It is likely, therefore, that Figure 2 understates the extent to which the Regional Accounts fail to capture the earnings outperformance by the South East in the 1980s.

Let us turn to the New Earnings Survey data. The NES is based on a sample of around 150,000 individuals compared with around 7,000 households for the FES, has a higher response rate and overlaps of around 75 per cent in individuals sampled year on year which reduces its variance further compared with the FES. However, it has the disadvantage for comparison with the Regional Accounts data of not being based on the place of residence and of being an

---

8 To be precise, the allocation across the regions to obtain Regional Accounts estimates is on the basis of each region's share in the sample wage and salary bill, that is, the product of average earnings per tax record times the number of tax records.
earnings per employment rather than an earnings per person measure (some people who have more than one employment could appear in the NES more than once).

Little has been written about the regional representativeness of the NES. However, by comparing regional shares of employment for different categories of workers for the NES with those for the Census of Employment, we can obtain some insights into this question. Figure 3 shows the South East share in employment in Great Britain of all full-time workers and of part-time women (who constitute the great majority of part-time workers). The figure shows that between 1979 and 1990, the NES under-estimated the share of the South East in full-time employment, and has under-estimated the share for part-time women ever since 1978. Since, as noted in section 2, the NES tends to undercount those employees with earnings below the National Insurance or Income Tax thresholds, one might have expected the share of the South East to be over-stated due to its higher average earnings. We cannot be certain why the surprising under-estimate of South East employment share occurs, but it is possible that the NES systematically undercounts employment in new businesses (particularly since as much as a quarter of the NES sample between the early 1980s and 1994 was derived directly from large organisations, see footnote 4 for details). Since we do not know why the under-counting occurs, we cannot be sure of the direction of any bias that may result in average earnings.

Though the comparison is not ideal (see further discussion in Section 3, below), we can compare the South East earnings premium in the NES measure of average earnings per employment with the Regional Accounts measure of earnings per employed person. To obtain earnings per employed person resident in the region from the Regional Accounts data, we need a count of employees resident in the region. Surprisingly, these data are less easily available than might be supposed. In the Labour Force Survey Historical Statistics, annual data from 1984 are

---

9 For example, the South East share of full-time workers was estimated at 33.7 per cent by the NES and at 35.1 per cent by the Census in 1985 (the year with the largest discrepancy), and the South East share of part-time women was estimated at 29.8 per cent by the NES and at 32.3 per cent by the Census in 1986 (the year with the largest discrepancy).
provided. For earlier years, we have two basic possibilities. One is to use unpublished data from the 1979, 1981 and 1983 Labour Force Surveys, these surveys not having taken place in 1980 and 1982. However, these data, unlike the 1984-96 data, have not been consistently matched to Census and other data sources to provide a strictly comparable series. Nevertheless, it is plausible that the South East to Great Britain employment ratio may be fairly reliable for these years.\(^\text{10}\) The result is shown in Figure 4 in the series marked RAa (which shows a continuous line from 1984 to 1995 and point observations for 1979, 1981 and 1983). For 1979-84, this shows an astonishing fall in the ratio of employment income per employee in the South East to the average for Great Britain. Not only does this contradict the NES data showing a relative rise in earnings per employment located in the South East, but other labour market indicators such as regional differentials in the unemployment rate, the rate of change of employment, redundancies and business failures. It is well known that the recession of the early 1980s struck traditional manufacturing and mining sectors, which are under-represented in the South East, particularly heavily.

It could be argued that the Labour Force Survey figures for 1979, 1981 and 1983 for relative employment levels in the South East are defective.\(^\text{11}\) Another possibility therefore is to ignore the 1979, 1981 and 1983 Labour Force Surveys and use a variant of the econometric model estimated by Cameron and Muellbauer (1998) on data for 1984-1995 to predict regional employment numbers before 1984. This model was developed in the context of explaining regional commuting rates and takes the following form. Let \(NR = \text{number employed in a region/number of employees resident in a region}\). We show that \(\ln NR\) defines the net

---

\(^{10}\) A slight variant of this is to use data on numbers in the regional labour force, see Employment Gazette, August 1986, which have been brought onto a consistent basis to adjust the regional employment figures which have not. However, this makes a difference of only around one half of one percent comparing 1979 with 1984.

\(^{11}\) In 1977, 1979 and 1981, the LFS data did not include people in employment training, or people in full-time education with jobs. We therefore adjusted the LFS data in these years to correct this discontinuity.
The commuting rate i.e., the rate of in-commuting minus the rate of out-commuting. Using the predicted ratio of employees resident in the South East to Great Britain, we obtain the line marked RAb in Figure 4. Although the fall is less pronounced, there is still a decline from 1979 to 1984 in average earnings in the South East relative to those in Great Britain. Furthermore, compared with 1979, even by 1987 there is little relative rise. However, from 1991 to 1994, the ratio of Regional Accounts earnings per resident employee for the South East relative to Great Britain rises quite strongly while the NES earnings per employment ratio suggests a temporary plateau has been achieved. Given the severity of the recession in the South East which, for the first time in the post-war era saw the average unemployment rate in London exceed that in Scotland and in some other regions, the NES profile looks, on the face of it, more plausible than the Regional Accounts profile.

Lastly in Figure 4, we have the Regional Accounts measure of income deflated by the number of employment documents estimated by the Inland Revenue, shown as RAc, which is effectively a measure of earnings per tax unit (the number of documents is only available from 1982). This measure is discussed further at the end of section 3, but it will suffice to say at this point that it behaves broadly similarly to NES earnings and suggests that the main problem with the Regional Accounts sampling method is that it got the regional proportions of tax units wrong.

To be precise, \( NR_i = \frac{NCE_i}{(1+PS_i)} NFLSi \), where for region i, NCEi is the number of jobs from the Census of Employment, PSi is the proportion of employees with second jobs and NLFSi is the number of employees resident in the region from the Labour Force Survey. We estimate a model for \( \ln NR_i - \ln NR_{gb} \) using regional data on earnings, unemployment rates, house prices and other variables. From this, we can predict \( \ln NR_i - \ln NR_{gb} \) for earlier years. Since \( \ln(NLFSi/NLFS_{gb}) = \ln(NCEi/NCE_{gb}) - \ln(NR_i/NR_{gb}) - \ln((1+PS_i)/(1+PS_{gb})) \), with predictions of \( \ln(NR_i/NR_{gb}) \) and \( \ln((1+PS_i)/(1+PS_{gb})) \), we can obtain good predictions of the place of residence count of employment for region i relative to Great Britain. As our paper demonstrates, the net commuting rate for the South East fluctuated by relatively small amounts from 1984 to 1995, and census data for 1981 and 1991 also show little change between those years, see below for further discussion. This means that the model, which relies heavily on the regional Census of Employment count to predict the regional place of residence count, should be reasonably accurate. Its standard error for the South East is 0.005.

Note that measure 'RAb' has been calculated on the basis that the proportion of second jobs in the South East relative to Great Britain was constant before 1984. This is likely to be a good
3. Possible Explanations for the Divergent Regional Accounts Data

We now discuss a number of possible explanations for the discrepancy between the Regional Accounts and the NES data: changes in commuting patterns; the increasing prevalence of part-time work; the move towards sub-contracted employment that is not covered by the PAYE scheme; changes in the relative fortunes of different kinds of workers; differences between social security and occupational pension contributions by region; and the problem that the Inland Revenue had in allocating employment records across the regions.

First, implicit in the above discussion is that it is unlikely that the discrepancy between the region of residence based Regional Accounts and the regional of employment based NES data could be due to changes in commuting patterns. To consider more formally what the relationship between the two should be, consider RER, the ratio of average earnings of those whose jobs are located in the South East to the average earnings of employed residents in the South East. The former consists of in-commuters and locally employed residents. The latter consists of out-commuters and locally employed residents. Thus

\[ RER = (\theta_1 \, w_1 + 1-\theta_1)/(\theta_2 \, w_2 + 1-\theta_2) \]

where \( \theta_1 \) is the ratio of in-commuters to the resident locally employed population and \( w_1 \) is the ratio of average earnings of in-commuters to average earnings of residents employed in the region, \( \theta_2 \) is the ratio of out-commuters to the resident employed population and \( w_2 \) is the ratio of their average earnings to the resident locally employed population.

Since \( \theta_1, \theta_1 \, w_1, \theta_2 \, w_2 \) are small, we can approximate as follows:

\[ \ln RER = \theta_1 \, (w_1 - 1) - \theta_2 (w_2 - 1) \]

Approximation since over the period 1984 to 1995, the proportions did not diverge by more than 0.5 per
Data which have become available from the LFS only in recent years on the relative earnings of commuters to those of non-commuters, show very similar ratios of around 1.6 for commuters into the South East as well as for commuters to other regions. Moreover, the limited data available suggest little movement in these ratios over time.

With \( w_1 \approx w_2 = 1.6 \), this suggests

\[
\ln \text{RER} \approx 0.6 (\theta_1 - \theta_2)
\]

In other words, the log ratio of place of employment to place of residence based average earnings for the South East should vary somewhat less than the net migration rate, assuming the relative earnings of commuters to non-commuters are fairly stable. The commuting model in Cameron and Muellbauer (1998) suggests that between 1979 and 1987 the net commuting rate for the South East increased by 2%. This can only account for a relatively small part, therefore, of the RA/NES average earnings discrepancy. For anyone who mistrusts either the LFS source of data for this model or the model itself, Census data suggest broadly similar conclusions. The most accurate figures on regional commuting come from the 1981 and 1991 Population Censuses. The rate of in-commuting into the South East of 1.8% in 1981 had risen to 2.3% in 1991 and the rate of out-commuting had risen from 0.7% in 1981 to 1.1% in 1999. Not only are these figures small in absolute terms, but they also show relatively little change.

Second, we can also dismiss the possibility that changing patterns of employment towards part-time work could help explain the discrepancies between the Regional Accounts and NES sources on regional earnings. As noted earlier, both sources rely on 1% samples of PAYE records which tend to omit earners, particularly part-time women, below the National Insurance Contributions floor. Total regional income from employment thus tends to omit these earnings and deflating by the LFS count of all employees, including part-timers, tends to underestimate average earnings. However, this should apply less to the South East, where average part-time
earnings are higher, than to Great Britain as a whole. If anything, this would indicate a small upward bias in the Regional Accounts estimate of the ratio of average employment income in the South East to that in Great Britain. Moreover, between 1979 and 1987, the proportion of part-time women in total employment rose around 2% less in the South East than in Great Britain from a common base of 17%. The upward bias in the earnings ratio should thus have increased slightly. This makes it even more difficult to understand why the South East earnings premium from the Regional Accounts sources rose so much less than from other sources.

Third, in recent years, the tax authorities have become a little concerned by the tendency of companies to shift employees previously on company PAYE schemes to subcontractors and personal service companies (where an employee leaves employment and sets up as a company where they can subcontract their services to an employer) where they are often not on PAYE. However, the comparison between Regional Accounts and NES sources for earnings should be largely unaffected by this as they are both from 1% samples of PAYE records, respectively of tax and national insurance.

Fourth, one can also dismiss changes in the regional differentials of different types of workers as an explanation. Figure 5 shows the rise in relative earnings per employment taken from the NES in the South East for a number of different kinds of worker: full-time men, full-time women, part-time women, full-time manuals, full-time non-manuals, and all workers. While the individual patterns vary, there is a surprisingly consistent rise in all the measures between 1979 and 1987 with changes of 5.7%, 4.7%, 3.8%, 3.8%, 3.4% and 6.7% respectively. The fact that the figure for all workers rises by more than its components suggests that employment growth in the South East has been slightly faster for better paid workers.

Fifth, it is clear that the distinction between income from employment, which includes employers’ national insurance and occupational pension contributions, and wages and salaries, which do not, also cannot explain the anomalous behaviour of the Regional Accounts income differentials. The data on wages and salaries are published only from 1984. Between 1984 and
1990, there are only small differences in the regional differentials from the two sources and after 1990 the two are virtually indistinguishable.

We come then to the final hypothesis: that there are problems in the methods of construction of the Regional Accounts income data. Unpublished data for 1982/83 to 1995/96 kindly made available to us by the Inland Revenue\textsuperscript{14} provide a possible clue. Between the 1982-3 and 1989-90 tax years, the Inland Revenue were unsuccessful in allocating around 12\% of United Kingdom tax records in their 1\% sample to a particular region. Beginning in 1990-91, the percentage unallocated to a region was progressively reduced so that by 1995-96, it had been reduced to only 1\%. Information on the percentage unallocated to a region for earlier years is not available. If the regional incidence of the unallocated tax records has varied significantly between 1979 and 1987, it could explain some of the apparent discrepancy between the Regional Accounts and NES regional earnings data.

An important piece of evidence comes from the ratio of the number of PAYE tax records allocated to the South East to the total number of allocated PAYE tax records for Great Britain. In Figure 6 this\textsuperscript{15} is plotted against the South East/ Great Britain ratio of the LFS count of resident employees and the Census of Employment count of jobs. It is remarkable that between 1982/3 and 1987/8, there is no rise in the relative number of tax records in the South East, while both the other sources show a substantial rise, around 40 per cent of the total 1979-1987 rise. This suggests an increased undercounting of earners in the South East in this period, which could therefore explain why the Regional Accounts measure of income from employment in the South East fails to rise in line with data from other sources. Furthermore, the sharp rise in the relative number of South East tax records in 1993/4 and 1994/5 coinciding with a sharp fall in the percentage of tax records not allocated to a region, contradicts the other two measures and suggests that the relative rise in the Regional Accounts measure of income from employment

\textsuperscript{14} We are extremely grateful to Jude Hillary in this regard.
and wages and salaries in the South East at this time may be spurious. Indeed, Figures 2 and 4 show just such a divergence in the Regional Accounts measure of the regional differential from other sources in 1993-4.

Together, the evidence is consistent with the hypothesis that in the 1979-87 period, there was a rise in the South East’s share in tax records that could not be allocated to a region and that this led to a serious underestimate of the rise in the South East’s share of national income from employment. It is not clear why the problem of the unallocated tax records occurred at this time, but it may have been influenced by the period of reorganisation and cutbacks in the government statistical agencies. In the 1990s, as increased efforts were made to correctly allocate tax records to the region of residence, this bias was corrected, leading to a spurious rise in the share of the South East. With under 1% of tax records unallocated since 1995/6, this source of bias has been removed. Indeed, Figure 4, shows only small discrepancies between the Regional Accounts measures ‘RAa’ and ‘RAb’ and NES sources of average earnings in 1995, of the order of magnitude one might expect, given commuting rates, between residence based and place of work based sources.

This leaves the final question of what the Inland Revenue and the ONS can do to correct this highly misleading pre-1995 historical record. Without resorting to the radical step of switching to an entirely different data source, such as the Family Expenditure Survey which has a relatively small sample and response rate problems, for pre 1995 data, much of the bias can be eliminated by switching to a per tax unit measure of earnings and scaling up these data to a regional level using Labour Force Survey data matched with regional population data. It is notable that between 1982/3-1990/1, the average earnings from tax units which could not be allocated to a region were between 80 and 85% of average earnings from all tax units. This suggests that the selectivity bias in computing average earnings per tax unit for the South East

---

15 We show the ratio for main employment documents. The ratio for all allocated employment documents moves very similarly.
and Great Britain resulting from the South East bias in unallocated tax units will be far less than the bias in the Regional Accounts data in Figure 4 measures a and b. For 1982 to 1995 we can compute such data and show them in measure c in Figure 4. Indeed, the pattern of movement is now broadly similar to the NES data on the South East earnings premium.

4. Consequences and Conclusions

In this paper we have compared Regional Accounts data on employment income and wages and salaries with other sources, in particular the Family Expenditure Survey and the New Earnings Survey. Especially between 1979 and 1987, the Regional Accounts source shows strikingly lower relative earnings growth in the South East than the other two sources. We considered five main hypotheses that might account for these differences, in particular the difference between place of residence and place of employment based data. At best, these can explain only a small part of the anomalous behaviour of the Regional Accounts data.

Instead, we provided evidence that problems in the method of construction have caused serious biases in pre-1995 Regional Accounts data. Between 1982 and 1990, the Inland Revenue were unable to locate the region of residence for around 12% of the 1% sample of tax records used to distribute national estimates of income from employment and wages and salaries across the regions. The data suggest that a disproportionate fraction of these unallocated records were in the South East, so resulting in an underestimate of earners in the South East. As the unallocated fraction was progressively reduced in the 1990s to under 1% in 1995, the bias was corrected. But in the process, the misleading impression was created of above average earnings growth in the South East at this time.

These biases have affected not only income from employment but also personal disposable income and gross domestic product, of which income from employment is a large component. Figure 7 shows South East to Great Britain ratios of income per head for income
from employment, personal disposable income and GDP, contrasted with an estimate of average earnings derived from the New Earnings Survey, also on a per capita basis.\(^{16}\)

As far as economic research is concerned, there are two important areas where it is likely that wrong conclusions may have been drawn as a result of these biases, namely, research on consumption and research on regional convergence. Data on personal disposable income for the United Kingdom from the Regional Accounts and published in Regional Trends, which are dominated by income from employment, have been used in a number of studies of the regional consumption function (for example, Green and Hadjimatheou, 1990, Bayoumi, 1993, and Carruth and Henley, 1993). As we have shown, these income data give a seriously misleading picture of changing regional income differentials. It seems likely therefore that the parameter estimates derived in these papers are seriously biased. As Figure 8 indicates, unlike the Regional Accounts data, the earnings data from the New Earnings Survey appear to reflect both the rise in the fortunes of the South East during the 1980s, and the stabilisation of that rise in the early 1990s when the South East was hit relatively hard by recession.

A possible consequence of the under-estimate of the 1980s rise in income in the South East in these studies, is that the effect of income on consumption may have been underestimated, while the effect of housing wealth, which rose disproportionately in the South East, may have been over-estimated. However, since the cited studies also suffer from a number of specification problems, such as the omission of financial wealth, of income expectations and of measures of income uncertainty, such as the change in the unemployment rate, further conclusions will have to await more systematic study.

Regional Accounts data have also commonly been used in studies of regional convergence that have looked at the United Kingdom (for example, see Quah, 1996, and Sala-i-

---

\(^{16}\) We adjust the NES earnings per employment in the South East to obtain an estimate of average earnings per worker resident in the South East (see Section 3). Multiplying this by the number of employees in the South East using the model discussed in footnote 12, and dividing by the population resident in the South East, gives per capita earnings.
Martin, 1996). For example, compare Sala-i-Martin (1996, Figure 9) which shows sigma-convergence in regional GDP per capita in the United Kingdom with Figure 8 of this paper which shows sigma-convergence (the standard deviation of log income) in various measures of Great Britain regional earnings. The three measures based upon the regional accounts all suggest a small rise (of around one per cent) in dispersion between 1979 and 1995 and look similar to the measure used by Sala-i-Martin, while the NES measure suggests a much larger increase in dispersion (of around four percentage points), close to doubling the later 1970s level of dispersion by the early 1990s. Evidence systematically contradicting the hypothesis of regional sigma-convergence casts doubt on the mechanisms for convergence and the universality of such convergence discussed by Sala-i-Martin (1996).

One implication of the downward bias in South East income levels in the 1980s is that income in poorer regions such as Wales and the North or North East regions will have been biased up. It will have misinformed public debate and public policy towards the English regions, Wales, Scotland and Northern Ireland. This should be of serious concern in a democracy.

Lastly, Regional Accounts data play an important rôle in the allocation of European Union Structural Funds. In order to qualify for Objective 1 status, an NUTS II region (in the case of the United Kingdom, typically smaller than standard regions and larger than counties) must have a level of GDP per capita less than 75 per cent of the EU average in a three year qualifying period which is typically four years before the relevant funding period. For the 2000-2006 period, the newly created Welsh NUTS II region of 'West Wales and the Valleys' has qualified for Objective 1 status. It is possible that if the biases in the Regional Accounts in the 1980s had not occurred, a Welsh region might have qualified for Structural Funds in either the 1989-1993 or 1994-1999 funding rounds. To indicate the funding levels at stake, Merseyside,

---

Barro and Sala-i-Martin (1995, p. 387, footnote 5) discuss the appropriateness of two different measures of US regional income, namely GDP per capita and Gross State Product (GSP) per capita. The latter is on a place of employment basis. Although Barro and Sala-i-Martin find that the different data do not make much difference to the rate of convergence, this approach has not been followed up in the subsequent literature.
which did qualify for Objective 1 status in 1994-1999, received around £85 per person per year. Had 'West Wales and the Valleys' been eligible for 1994-1999, with a population of approximately 1.5m, it might have expected EU funding to the tune of around £130m.

References


Notes: N A : N E S is the log ratio of National Accounts personal sector income from employment per worker to New Earnings Survey gross weekly earnings per employment (not affected by absence). N A p h : F E S (M A 3) is the log ratio of National Accounts personal sector income from employment per household to a three-year moving average of Family Expenditure Survey gross normal weekly household income from wages and salaries, both deflated by the consumer expenditure deflator. N A : A E I is the log ratio of National Accounts personal sector income from employment per worker to the average earnings (Great Britain) index (whole economy).

Notes: As Figure 1a but using National Accounts personal sector wages and salaries instead of income from employment. Income from employment includes employers' contributions to Social Security and other contributions, while wages and salaries do not.
Notes: The log ratio of South East to Great Britain for Regional Accounts personal sector income from employment per household and a three-year moving average of Family Expenditure Survey gross normal weekly household income from wages and salaries, deflated by the consumer expenditure deflator.

Figure 3

South East Share of Employment of Full-Time Workers and Part-Time Women based on NES and Census of Employment Data

Notes: The Share of the South East in Great Britain employment of Full-Time Workers and Part-Time Women according to the NES and the Census of Employment.
Notes: The log ratio of South East to Great Britain for New Earnings Survey gross weekly earnings per employment (not affected by absence), indicated by 'NES', as well as four other measures with the Regional Accounts income from employment as the numerator. The denominators are, respectively, Labour Force Survey estimates of employment, indicated by 'RAa'; estimates of employment based upon the model of Cameron and Muellbauer (1998), indicated by 'RAb'; and the number of main employment documents as supplied by the Inland Revenue, indicated by 'RAc'.

Figure 4
Log Ratios of South East to GB earnings for Regional Accounts and New Earnings Survey

Notes: The log ratio of South East to Great Britain earnings for various categories of NES workers, namely, full-time men (MFT), full-time women (FFT), part-time women (FPT), all workers (Total), full-time manuals (Manual FT), and full-time non-manuals (Non-Manual FT).
Notes: The log ratio of South East to Great Britain employment for various measures of employment, namely, the Labour Force Survey, the number of main employment documents (Inland Revenue), and the Census of Employment.

Figure 6
Ratios of Employment in South East to GB for different measures of employment

Notes: The log ratio of South East to Great Britain income for different measures of earnings, namely, Regional Accounts income from employment per capita, personal disposable income per capita, GDP per capita, and New Earnings Survey per capita (see footnote 16).
Notes: The standard deviation of log income for different measures of earnings, namely, Regional Accounts income from employment per capita, personal disposable income per capita, GDP per capita, and New Earnings Survey per employment.