

Understanding and improving data quality relating to low-income households

David Johnson and Rosanna Scutella *

Melbourne Institute of Applied Economic and Social Research

Abstract

Concerns have been expressed about the reliability of income and expenditure data from Australian Bureau of Statistics surveys for those on low incomes. This paper analyses the factors behind the apparently low incomes recorded by many Australian households, in particular, in major ABS data collections, and the implications of these for income distribution and 'poverty' analysis. The paper seeks to identify the extent to which these arise from the concepts of income used in these surveys, which may result in the data being inappropriate for other analysis; and the degree to which it can be ascribed to inadequate reporting of incomes. The paper draws some conclusions about how the existence of such problems should be treated in income distribution and related poverty analysis, and whether any specific actions should be taken by analysts to derive data appropriate for the purposes of their analysis. It also considers some methodologies that can be adopted to reduce the impact of any identified problems in future data collections.

* We acknowledge the help of Tim Watts in undertaking some early work on this project

Table of contents

1. Introduction	3
1.1 Motivation	3
1.2 Low income households	4
1.3 In this report	5
2. Reliability of measurement.....	6
2.1 The policy context	6
2.2 Errors in sampling	9
2.3 Grossing up income surveys.....	10
2.4 Comparisons of annual and current (weekly) income	12
3. Comparing income and expenditure from surveys.....	12
3.1 The HES	13
3.2 Households with zero or negative income.....	14
3.3 Comparing households in the bottom equivalised income and expenditure quintiles.....	14
3.4 Households with low income but not low expenditure	17
3.5 Households with low expenditure but not low incomes	20
3.6 Households with both low income and low expenditure.....	20
3.7 Summary	20
4. Government transfers in administrative data and income surveys	21
4.1 Detailed explanations of the data sources.....	21
4.2 Recipient numbers	22
4.3 Total income from benefits.....	26
4.4 Average incomes	28
4.5 Income distributions of observed benefit population	31
4.6 Key points.....	34
5. Reported vs imputed eligibility levels for income support recipients by payment type.....	35
5.1 Recipient numbers	35
5.2 Entitlement levels	38
5.2 Income distribution.....	39
6. Concluding comments and implications.....	41
6.1 Are the data relating to low incomes plausible?.....	41
6.2 Overall interpretation of results	43
6.3 Strategies for dealing with data quality concerns	44
6.4 Further work	46
References	47
Appendices	49

1. Introduction

Social policy is focussed on low-income families. While there is much anecdotal evidence about the experience of the poor, policy development requires accurate measurement of their circumstance. Measurement generally relies on survey and census data. Until recently in Australia, the key sources of data were produced by the Australian Bureau of Statistics (ABS).¹ The primary collections were the income surveys, the expenditure surveys and the census. The availability of unit records of respondents in the surveys provided researchers with an extremely valuable means of calculating poverty, inequality and social welfare. However along with the burgeoning of studies utilising the data were concerns about their reliability. Concerns have been particularly expressed about the reliability of income and expenditure data for those on low incomes.

1.1 Motivation

Income is one of the most frequently used variables in social analysis. ABS surveys of family income are the major source of data for calculations of poverty, inequality and social welfare. Analysis of income distribution and income poverty focus on the pattern and level of low income households. To the extent that these reported incomes do not provide an accurate picture of the provision of income support or the resources available to households, analyses based on them will be flawed.

Two sources commonly used by researchers for such studies are the ABS income surveys, which we shall call the IDS until 1990 and the SIHC from 1994; and expenditure surveys called HES.² The ABS has made available confidentialised unit record files for seven of the eight IDS/ SIHC, spanning the period 1981/2 to 1997/8, and all five of the HES. Unfortunately there have been significant changes in the sampling frame, in the questions asked and in the variables recorded for respondents in the surveys and there is a lot of incompatibility between them. Nevertheless the information from these surveys provides important new insights into the actual incomes flows for low-income households, although there are limits to its robustness.

These two sets of surveys show:

- a proportion of Australian households recording zero or negative incomes (for example 1.2% of households in the HES);
- other households with very (implausibly) low incomes;
- many households, apparently in receipt of income support, with incomes less than would be expected given rates of payment;
- apparent under-reporting of receipt of income support; and
- implausibly higher marginal propensities to consume.

Analysis of surveys such as HES suggest that there is a significant discrepancy between reported incomes and expenditure. The consumption levels and patterns of many of these low-income households and the consequential incidence of 'financial stress' are not easily reconciled with their income level.

¹ In 2001 the Melbourne Institute gathered the responses of around 8000 households in the first wave of a longitudinal panel survey of Australian households. The survey, known as the HILDA (Household, Income and Labour Dynamics of Australia) survey will provide an invaluable alternative data source once a reasonable number of waves (years of response) become available. Burkhauser and Smeeding (2002) have recently described the value of such data sources. For more on HILDA see www.melbourneinstitute.com

² The income surveys we refer to are the Income Distribution Surveys prior to, and including, 1990 and the Survey of Income and Housing Costs from 1994 to present. The expenditure survey is the Household Expenditure Survey.

Where data collections such as the IDS/ SIHC collect information on both current and previous full-year income a diverse pattern of discordance exists. If there is confusion between households with genuinely low incomes and those with apparently low incomes due to reporting, or the conceptual basis of the income variable being used (for example the treatment of business activities in economic concepts of income), there are two risks: one, the risk of misdirection of social policy and therefore wasted government resources, and two, the risk that seriously impoverished households are not identified or assisted.

Where data does not accurately reflect the circumstances of households there is the long-term danger that these data limitations will serve to reduce the overall credibility of social policy research.

1.2 Low income households

Understanding the nature of low income and /or expenditure groups is essential in the consideration of policy reform. For instance, contrary to the experience of many recent occasions when government has introduced new economic policy, in debate about the introduction of reforms to the tax system discussion centred on its effects on particular groups in the community. In particular there was concern about the extent to which particular groups would suffer from the introduction of the reforms and the adequacy of the compensation to be made available to them. Ensuing research aimed to identify the effect of the changes on particular groups in the community. Unfortunately the identification of particular groups in the data was not always as clear as might have been hoped. In particular it was not clear that the lowest income decile identified from national survey data matched the classic groups of poor that are the prime subject of concern. In terms of policy, the “really” poor – those who are chronically poor are, in the main, not in the lowest decile, but found in the next lowest decile; and they are the ones that need to be targeted in social transfer schemes.

Available evidence suggests that the market economy is developing in such a way as to increase the gap between households on high and low income (Harding and Greenwell, 2002; Johnson and Wilkins 2002). An important role of government has been to mitigate this increase to maintain the degree of dispersion in income after taxes and transfers and after provision of non-cash benefits (see Harding, 1997; Johnson et al, 1995, 1998). Thus there is a role for government in providing compensation for major changes in the economy, which are thought to impact unfairly on some groups.

In order to compensate the target needs to be explicitly identified is. This turns out to be a very tricky issue. There is evidence about potential client groups from the articulation of concerns through lobby groups –the elderly, unemployed, sole-parent pensions, low-wage income earners, large families etc. However while the target groups may be readily identified from lobbying activity and from the case studies detailed by the media, describing the circumstance of such groups cannot be assured from the likely biased presentations of the groups or the highly stylised presentations in the media. Fortunately, the availability of national income and expenditure surveys affords the opportunity to identify the circumstance of very many such groups and of their relative number in the population. But distinguishing the real disadvantaged from the apparently disadvantaged from these surveys is not straightforward. There are both conceptual issues and practical issues.

From a conceptual point of view there is debate about the appropriate measure of wellbeing, for instance whether income or expenditure provides the best medium in which to measure of the welfare of individual households. There is a very large literature which argues that much broader measures of welfare incorporating education, health freedoms, social networks are needed to identify wellbeing but in the absence of data needed for broad measures we put this issue to one side. Income is the frequently used unit, for instance poverty lines are expressed relative to income, national wellbeing is often gauged by gross domestic product per capita, and so on. However many researchers (for example Travers and Richardson, 1993) have pointed out that while income provides an indication of resources available to households it is not complete and in any case resources do not indicate how households actually live. Resources may be used wastefully or wellbeing may be supplemented by consumption

provided by government (for instance through non-cash benefits such as health and education). They argue that for household wellbeing a measure of consumption is better.³

There are also practical problems with the identification of groups in the community. The most obvious practical problem is that we don't know the circumstances of the groups and we have to rely on information about samples of the population and our judgement to evaluate their situation. There are many problems in the process of gathering information and critics of targeting compensation such as the federal treasury have pointed out these problems (see Carnaghan, 1998). Nevertheless the national surveys of income and of expenditure provide invaluable information about the circumstance of individual households and of groups of households and indeed there are few alternative information sources.

1.3 In this report

The job of this paper is to compare the various sources of income (mainly) and expenditure (to a lesser extent) and to see if any patterns emerge. The paper has three objectives:

- to better understand the factors behind the apparently low incomes recorded by many Australian households, in particular, in major ABS data collections, and the implications of these for income distribution and 'poverty' analysis;
- to identify the extent to which these arise from the concepts of income used in these surveys, which may result in the data being inappropriate for other analysis; and the degree to which it can be ascribed to inadequate reporting of incomes; and
- to provide advice on the magnitude of any problem which may exist; how the existence of such problems should be treated in income distribution and related poverty analysis, including any actions which should be taken by analysts to derive data appropriate for the purposes of their analysis; and consider methodologies which can be adopted to reduce the impact of any identified problems in future data collections.

We shall consider the following sets of data:

- weekly and annual income from income surveys;
- income data from expenditure and income surveys;
- income data for recipients of government transfers from administrative records;
- expected income data for recipients of government transfers from application of rules and regulations.

In the next section we discuss the problem of the reliability of low-income data in a broader context.

Sections 3, 4 and 5 contain the novel contribution of the report. We provide a systematic comparison of a variety of data sources in order to explore issues concerning the nature of measurement of low income. In section 3 we compare income measured in the ABS income and expenditure surveys.

In section 4 we focus on a comparison between administrative data sources (the longitudinal data set of recipients of government cash transfers or LDS⁴) and the income surveys. In section 5 we compare income as recorded in the administrative database and the surveys with the expected income according to entitlement as portrayed in our MITTS model.

In the final section we review our findings and suggest some strategies for researchers and policy makers to deal with concerns about data reliability.

³ The Poverty Group 2001, p7 support this view, particularly in relation to developing countries

⁴ These are the confidential client records from the Centrelink agency. While Centrelink gathers the information from recipients the data have been supplied to us by the governments policy making agency, the Department of Family and Community Services.

2. Reliability of measurement

The motivating concern is the reliability of measurement, particularly as it applies to low-income groups. However this motivation needs to be tempered by the context in which measurement is made and used. Primarily measurement is required to guide policy. But policy is subject to a range of factors of which data quality is just one. While the focus of this report is the issue of data quality it is useful to first note all the reasons for puzzling, unclear or conflicting measurement.⁵

2.1 The policy context

Concern about data quality emanates primarily because of the limitations that are placed on policy development. Policy development relies on accurate and timely information and doubts about quality limit the depiction of existing situations. We consider two issues, the form in which information is presented to policy makers and the nature of the information presented.

The form of information

In providing information to policy makers there is a trade-off between simplicity and accuracy. Characterising a complex situation in terms of simple examples provides a powerful tool for conceptualising policy solutions but runs the risk of obscuring more complicated underlying issues.

Often debate is couched about the circumstance of particular groups. While the use of a representative of a particular class may often be helpful in defining a problem it also needs to be recognised that there is often very wide variation within groups and the use of representative data may unwisely and erroneously simplify more complicated situations.

Use of representative groups

A concern in relation to the interpretation of results refers to the frequent use of representative or average observations to depict the circumstance of groups. For instance research is often augmented by documentation of the situation for particular groups of people who may be considered disadvantaged. This is a very common ploy of the media, to highlight particular issues and to make particular situations more meaningful for their readers, listeners and watchers. For instance these groups might be:

- Families with unemployed breadwinners;
- Families of sole parent pensioners;
- Old age pensioners;
- Families dependent on heads on disability pensions;
- Families and individuals belonging to particular ethnic groups including indigenous people;
- Families and single persons dependent on low-wage income; and
- Other low-income groups.

Descriptions of these groups would characterise them as being typically in some financial and social distress; being dependent on government transfer payments, experiencing difficulties in meeting payments for necessities, going without normal activities and being unable to participate in what might be regarded as conventional social activity.

Frequently policy debate is focussed on the circumstance of representatives of such groups. Understandably there are lobby groups for each who may forcefully represent the interests of their group. However in considering groups the researcher and the policy maker has to be careful that stylised depictions of such groups, even averages, may not represent many

⁵ Atkinson, Brandolini and Smeeding (2001) discuss many of the relevant issues in relation to the production of time series data on income distribution.

individuals at all. It is often the case that variation within such groups is very great, frequently dominating variation between groups. Great care must be taken when drawing inferences from the alleged circumstance of such cases.

The nature of information

Frequently data quality is just one of a number of possible causes of counter-intuitive or unpredictable or conflicting differences in measurement. Alleged differences in measurement in relation to inequality and poverty may be the result of a number of circumstances. Differences may occur in interpretation, in definition and in data. At the broadest level people may simply interpret the same results in different ways. At the next level people may produce different results with the same data because they use different definitions. Finally people may use different data which measure the same concept but with different results.

Problems of interpretation

Differences in interpretation arise because individuals place objective information in different perspectives. Perspectives are based on the experiences and world view of the individual and there is no invalid perspective. Research has nothing to say about differences in interpretation. Such differences do not depend on objective information though they may be subject to logical analysis.

Problems of definition

A second major cause of alleged differences in measurement arise because of differences in definition. People use the same terminology to refer to different, sometimes conflicting concepts. First there are differences in the meaning of an idea. Consider poverty. There is an important and ongoing debate about whether poverty is defined in a relative or an absolute way or in some combination of both. Absolute measurements are invariant to the circumstance of others whereas relative measures are conditioned by the circumstance of others. Poverty is typically defined through comparison with norms in the population, but what those norms are is open to interpretation. Whatever choice is made (and it may not be explicit) will lead to a different methodology and different results.

Second, different measures may be attached to the same idea. Consider wellbeing. The ultimate aim of society is to improve the wellbeing of society. There are a wide array of definitions used to proxy wellbeing in relation to distributional and equity analysis. Wellbeing is frequently measured by disposable income adjusted for family size and composition. But in general, income is seen as an indirect measure of well being, providing the opportunity for income recipient groups to achieve a standard of living. Total expenditure per head is argued to be a better measure of the actual living circumstance, allowing income to be smoothed over time. Again the measured result will depend on the choice made here.

Third, within income no single definition is commonly accepted and income might be measured at many levels, wage and salary income; private or market income, gross income, disposable income, social wage income. Private income includes wages, salaries, income from self-employment and income from rents, interest, superannuation and dividends. Gross income is private income plus transfer payments received from government. Government transfers make up a significant proportion of many low-income households and include unemployment benefits, training wage, old age pension, disability pensions, parenting payments, rent assistance and family payments. Over the entire population transfers average about 20 percent of income whereas for the elderly, the unemployed and single parents it may be closer to 100 percent. Disposable income is gross income less personal income tax.⁶ Behaviour is determined by income after tax, so that is generally the most relevant concept for the development of policy.

⁶ While respondents usually know their gross income they are less likely to provide reliable information about tax so in many surveys, personal income tax is estimated according to knowledge of tax schedules and observed aggregate behaviour of tax entities.

Fourth, an important issue is the extent to which income (or total expenditure) is adjusted for differences in need, most frequently depicted as the size and composition of the unit of observation. Income is seen as a proxy indicator of potential to achieve a certain level of wellbeing. It is not the *actual* level achieved. Adjustment for needs is brought about by converting cash income to equivalent income where there is provision for the number of family members, age, sex and work force status using “equivalence scales”. There are limitations of this process– they equate the needs of families of a particular type at a particular level or standard of living. Other criticisms are that equivalence scales over-simplify and over-generalise much more complex situations. However there are also conceptual concerns. In conventional uses of equivalence scales, a large number of children counts for a lower standard of living. But revealed preference arguments suggest that the choice of families to have children is because they provide higher utility, and presumably a higher standard of living.

Fifth, there are other sources of utility not included in disposable income. Non-cash benefits – government provision of health and education services – may be quite (and variously) important at some times and in some jurisdictions (countries). The flow of services from assets is also not considered in the most widely used definitions. Wealth may be important providing “insurance” in times of stress and loss of income, or access to loans to cover periods of unemployment. Important forms of assets (such as homes) may not necessarily be (easily) marketable or liquid. Net worth (assets less liabilities) is more unequally distributed than income, both are positively correlated, and wealth tends to increase disproportionately with age, so if it is factored in will increase well being estimates for the aged in comparison to young people. Full income includes the value of services flowing from durable goods (ie the imputed rent from owned housing stock), the value of in-kind transfers, non-cash benefits and the value of leisure. There is a trade-off between leisure and work, so that there will be divergence between gross and full income.

Sixth, there are different valid observational units of analysis. The unit of observation may be the individual, income unit, household or family. Choices of particular groups to be the point of comparison may have a significant impact on results. Choice of group and size of group can be made to either highlight or suppress particular outcomes.

Finally there is the issue of time. Frequently surveys capture measures of wellbeing at a point in time whereas wellbeing is experienced over time. If mobility is high, poverty apparent from cross-section surveys may not be a problem when measured over the lifetime. Consideration of the pattern of poverty over time raises other issues – how is poverty over an extended period (say 2-3 years) of low income or expenditure rated relative to poverty over a short period. Measurement of poverty in cross-sections may be compromised by not taking into account a longer period of analysis. For instance negative income may occur when a small business operator faces temporary periods of operating losses.

Problems with data

When allowance is made for differences in interpretation and proper account is taken of definitional differences residual differences may be due to data quality. There are two classes of error associated with data:

- Non-sampling error; and
- Sampling error.

Non-sampling error is concerned with the problems that arise when the sample is not representative of the population and weights attached to observations do not reflect the population or when respondents give incorrect responses for one reason or another.⁷ Some examples of non-sampling error include:

- Incorrect reporting of respondents answers through mistakes by the interviewer;
- Respondents intentionally giving incorrect answers;

⁷ For more on these issues see Moser and Kalton (1971)

- Respondents unintentionally giving incorrect answers because they forget (this problem is magnified the greater the time period over which they are asked to recall information);
- Respondents unintentionally giving incorrect answers because they do not understand the question (for instance a common mistake is for the disability support pension to be reported as sickness allowance);
- Inappropriate editing of data and imputation for missing observations;
- Errors arising in the transfer of data from questionnaires to the data made available to researchers;
- Changes over time in the way data is collected and processed, for instance changing the medium of collection from face to face to telephone will have an effect on the way questions are perceived and on the way responses are transmitted.

Sampling errors refer to the expected reliability of the data based on the number of observations. The greater the number of observations the lower will be the errors associated with summary measures derived from the observations. Similarly more complex derived variables may have greater sampling error (for instance errors may be magnified in derived variables that use more than one observation).

Sampling error is likely to lead to volatility in a time series, however many causes of non-sampling errors may remain constant over time. For example if the phrasing of a question is unchanged the extent to which misinterpretation occurs is likely to be constant over successive surveys.

As mentioned earlier, in Australia at present, the main sources of information on income and expenditure used in studies of inequality and/or poverty are the ABS income and expenditure surveys. However, questions have been raised as to the reliability of this data at a point in time, with a particular focus on the comparability of surveys across time. In an ongoing project with the SPRC, the ABS is delving into these issues in more detail. One outcome of this is an undertaking by the ABS to examine and update the survey data across the available set of years to make them comparable over time and thus suitable for any analysis of trends in income or expenditure distributions across time (see ABS, 2002). According to ABS 2002, the likely revisions to the survey data are not likely to make any significant differences to aggregate measures based on the general population, such as Gini coefficients across the sample as a whole. However, as subgroups are selected, and these samples become smaller, it is likely that the differences may be substantial.

2.2 Errors in sampling

When examining data sets like those addressed in this report, close attention must be paid to the degree of sampling error present in the numbers. Certain movements in the sample data over time may in fact not be reflective of trends in the actual population being examined: shifts over time may simply be a product of the small size of the sample rather than indicative of any important underlying trend in fundamentals. Confidence interval analysis is helpful determining whether the variations are significant statistically or simply represent movement within the standard error range for this particular sample from this particular population.

The ABS employs the standard error (SE) as a means of measuring the likely difference between an estimate derived from sample survey data and the reality if the entire population had been surveyed.⁸ As an indication of the size of the SE, there are about two chances in three that a sample estimate would differ by less than one SE from the figure that would have been obtained if the entire population had been surveyed, and about 19 chances in 20 that the difference will be less than 2 SEs. This relationship can also be expressed by using the Relative

⁸ Most ABS publications reporting income survey statistics carry an appendix providing detail about the calculation of standard errors. The discussion here is taken from the Appendix 3 ABS (2001).

Standard Error (RSE), which indicates the extent of the possible difference between the sample estimate and reality as a percentage of the estimate.

Generally the larger the size of the sample, the more reliable it is. With larger samples the size of the SE rises, the RSE (which is expressed relative to the total population) falls and greater confidence may be placed in the significance of trends. It is difficult to make useful conclusions from estimates generated from small sample sizes because they are subject to very high RSEs. Any estimate with an RSE of greater than 25% must be treated with some caution. Estimates RSEs greater than 50% are of only negligible value.

Knowing the size of RSEs is also a useful discipline when thinking about what conclusions may be derived from say movements over time in some variable of interest. If it were expected that the true movements over time in some variable were going to be relatively small, say less than 5 per cent, then to draw meaningful conclusions about trends you would need to have RSEs much lower than 5 per cent.

Strategies for dealing with data fluctuation

For every data fluctuation evident in survey data, it is necessary to ask whether the fluctuation is population driven or survey driven. The concern on some occasions is that it is not always possible to conclusively determine which of these factors is driving the fluctuation. In the case of some quite small estimates of sub-groups within the dataset it may be appropriate to apply particular subject matter knowledge when interpreting specific cases.

Given the existence of fluctuations, the following strategies may be helpful in utilising survey data:

- Calculate RSE to see if sample variability (linked to sample size) is an issue for their particular population sub-group;
- Use confidence interval analysis where RSEs are acceptable (i.e. sample size is not a big issue), in order to determine if the potential fluctuation is statistically significant;
- Search for outlier records that may be contributing to the fluctuation and to adjust individual records on an ad-hoc basis to minimize fluctuations; and
- Use ratio estimates rather than estimates of levels.

2.3 Grossing up income surveys

Since the 1950s all modern economies have developed systems of national accounts, which draw on a very wide range of data (including surveys) and aim to represent the value of aggregate production. The availability of national accounts provides a point of comparison for household data. One check that is often made of survey data is to compare the household level data, aggregated in an appropriate way, to the estimates in the national accounts of the corresponding aggregate measure.

Such comparisons have led to some concern about the accuracy of income from cross-sectional surveys. A recent study by researchers at the World Bank in 2001 noted that,

“Better methods for handling income surveys could also help to resolve the worrying discrepancies between survey-based measures of average economic welfare and National Accounts. Income surveys are more troubling from this point of view. There is only a 5% difference in the aggregate consumption estimates from surveys and National Accounts, but aggregate household income for surveys is on average 25% below the private consumption subcomponent of national accounts.” (The Poverty Group, 2001, p.7)

The World Bank report also states that, “The case is compelling on both conceptual and measurement grounds for the view that consumption expenditure is a better welfare metric than income in developing country settings.” (p.7). However the World Bank study focused on poor and undeveloped countries where underreporting of income and barter may be more important.

However in general the reliability of household data does not seem to have been a major concern of overseas researchers. The Luxembourg Income Study (LIS) is a non-profit organisation promoting cross-national research on poverty, inequality and social policy by collecting household data on income and demographic characteristics. LIS (2002) is a website for the Luxembourg Income Study, and describes a facility that presents income and expenditure data sets from 25 or so countries. These datasets are adapted in order to improve cross-national comparability. LIS is sponsored by organisations from each country in many cases the key national government statistical agency. Sponsoring agencies provide comparisons of income grossed up from surveys and national accounts data for surveys. Since its inception 20 or so years ago several hundreds of studies have been undertaken with the LIS datasets and are listed on their website. In the many discussion papers and reports published as a result of using the LIS few have directly emphasised reliability of the data and data quality matters. Behrendt (2000) is an exception. She notes that means tested benefits tend to be under-reported in income surveys. A working group instigated by the Australian Bureau of Statistics reported on many issues including the responsibilities of primary data providers to give guidance on data quality and limitations (see The Canberra Group, 2001).

Nevertheless the problem of conflicting results between surveys has been noted elsewhere. In the UK there has been concern about conflicting results between income and expenditure surveys. Frosztega et al (2000) detail comparison of the UK Family Expenditure Survey (FES) and the Family Resources Survey (FRS an income survey). Nordberg et al (1996) studied the effects of using interview versus register data in income distribution analysis and Moore et al (2002) review measurement errors in surveys in the US by comparing survey results with administrative information. They find that in US surveys respondents do tend to under-report transfer and self-employment income but that there are many reasons for this, including definitional problems, recall and salience issues and confusion as well as the much-suspected matter of data sensitivity. They also find that measurement of wage and salary incomes are accurate.

Australian national accounts

In Australia the Social Policy Research Centre has made comparisons between income surveys and national accounts estimates of household disposable income (SPRC, 2002). Once differences in definition are taken into account the SPRC found that the surveys consistently account for around 90% of the estimated aggregate national accounts household disposable income.

Bradbury (2002a and 2002b) has grossed up income variables in the 1982 and 1986 income surveys and compared them to estimates of household income from the Australian National Accounts (ANA). Kroon and McDonald (1998) have similarly grossed up income for the SIHC 1994-5 income survey and ABS (2002) report the grossed up income for the 1996-7 SIHC survey. In these two studies the raw grossed up aggregates for government benefit payments were further adjusted to take account of known differences in target populations, concepts and timing. Both found that between 60 and 70% of all differences in SIHC and ANA are due to conceptual differences. Other results were (grossed up SIHC as a percentage of ANA):

- Wages and salaries: SIHC accounts for 92% in 1981-2⁹; 101% in 1985-6, 104% in 1994-5 and 96% in 1996-7.
- Personal benefit payments to residents: SIHC accounts for 75% in 1981-2, 66% in 1985-6, 64% in 1994-5 and 67% in 1996-7. When adjusted the income surveys could explain 87% of transfers in 1994-5 and 91% in 1996-7.
- Household income: SIHC equal to 83% in 1981-2, 82% in 1985-6, 91% in 1996-7, and 89% in 1997-8.

⁹ Bruce Bradbury in a personal communication has pointed out that this low figure is largely the result of definitional differences (such as the non inclusion of owner-employees in aggregated wage and salaries).

In general there appears to be a substantial level of under-reporting of income among recipients of government benefits and pensions. Even when differences in target population, concepts and timing are brought to account their remains a ten percent shortfall. If this shortfall is constant in successive surveys then estimates of change over time will be valid. If however there is some temporal change in the extent of under-reporting the estimates of trends in poverty and inequality will be biased.¹⁰

2.4 Comparisons of annual and current (weekly) income

Another potential source of variation in data quality derives from the differences between current income measured on a weekly basis and annual income. The Social Policy Research Centre (SPRC, 2001b, unpublished) embarked upon a preliminary analysis that sought to compare how the differences between annual and weekly income data leads to different estimates of the rate of child poverty. Estimates of poverty derived from annual income tend to be much larger than those derived from current income in the SIHC.

The approach of the SPRC paper was to directly compare cohorts in annual income data and the current income data that would be expected to be similar. It used SIHC data for 1995-96 and 1996-97. It then focused upon the variations across the two data sets and sought to explain the source of the variations.

The SPRC found that the higher rate of poverty derived from the annual income data appeared to stem from two factors: One, median annual income is higher than median (annualised) current income thus implying a higher relative annual poverty line. Two, a greater proportion of elderly and lone parent have annual incomes below base pension rates. The SPRC reported that further research into the differences between average annual and current incomes which included decompositions by income and family type and a comparison of external data on income growth trends between the annual and current periods of the surveys would be worthwhile.

3. Comparing income and expenditure from surveys

A number of studies have utilised data from different surveys to compare income from different sources and to compare income and expenditure distributions. SPRC (2001a) compare income for the 1993-4 year from the 1994-5 SIHC with income for the same year from the 1993-4 HES. Harding and Greenwell (2002) compare income and expenditure inequality in Australia over the 1980s and 1990s. They compared income inequality across expenditure and income surveys, and then compare expenditure inequality with income inequality. They showed that income inequality rose in the 1990s using both income measures from both HES and IDS/ SIHC, although this result was not as strong as using the IDS. The results differ somewhat across surveys if you look at different points in the income distribution. For instance, relative to the expenditure survey, lower income households in the IDS/ SIHC fared better, middle-income households fared worse, and the high-income earners did not fare as well, also inequality did not change from 1994/95 onwards.

Expenditure inequality is not as great as income inequality (Barrett, Crossley and Worswick, 2000; Blacklow and Ray, 2000; Harding and Greenwell, 2002). Trends in expenditure inequality also differ from trends in income inequality and there was no clear increase in inequality using expenditure. Many argue that using expenditure to measure inequality is

¹⁰ Often the current poverty rates are reported and discussed in media. However debate about rates of poverty is often fairly meaningless. Since there is no generally accepted reference poverty line, nor are there generally accepted equivalence scales the choice of reference poverty line and equivalence scale is entirely at the discretion of the researcher. Effectively the researcher picks the poverty rate. In this context discussion of what is the 'correct' current poverty rate is simply a matter of opinion, Even so measures of change over time are of great interest, and for a wide range of choices of poverty line and equivalence scale will be of considerable policy interest. The choices will affect the results and must be transparent and explicit.

more appropriate than using income as transitory income fluctuations are removed with consumer expenditure used as a proxy of permanent income.

In this section we compare income and expenditure data using the 1998/99 HES, which includes measures of both income and expenditure. Another comparison could be made with income from the 1997/98 SIHC, however as the time periods differ over the two surveys the comparison would not be consistent. We could index the measures to be comparable, but this would introduce a further level of uncertainty that we do not feel is warranted.

3.1 The HES

A test for evaluating allegedly low-income households is to compare them relative to informal expectations about their nature. Does the picture of the low-income household in the surveys conform to the conventional picture of the needy? We use unit records of the ABS 1998-99 HES to understand and explain 'interesting' trends/characteristics as revealed by the statistical profile. For example, by examining all variables for a selection of the households citing zero income we may see whether we are dealing with the deeply impoverished, people making transitions in the labour market, self-funded retirees who may be 'asset-rich but income-poor' or new school leavers. It will also be interesting to look at the characteristics of the self-employed, in particular the relationship between their income and expenditure patterns, given that the self-employed constitute 30% of the households recording zero or negative income.

The latest HES provides information about the income and expenditure of a sample of 6892 Australian households and this information may be used to distinguish households by type (i.e. demographic or labour-force characteristics), by income and by expenditure. Using the raw (unequalised) data at the household level, we will begin by looking at the lowest income and expenditure quintiles as defined by the HES data and consider whether the characteristics of these households (or what may be regarded as factors causing, or being caused by, poverty) reflect the characteristics of the poor when defined by welfare and community agencies. We shall also consider other ways of defining the groups of interest. Note that expenditure on goods and services is used in this analysis less expenditure on income tax, thus omitting capital expenditure items. Also, the results are weighted to represent population estimates using the ABS survey weights provided.

To find out more about low-income households with high marginal propensities to consume we explore the characteristics of households in the bottom quintile defined by both expenditure and disposable income (hereafter income for short). We omit those with zero or negative disposable income as we do not believe such levels of income to be sustainable in the long term. We show the statistical profile for this omitted group in the first column of Table 1. As can be seen it is much different to the characteristics of the low-income population with positive incomes.

We do not expect all households with low incomes to be those with low expenditure as expenditure is used to proxy permanent income. In this way transitory fluctuations in income are removed and we pick up the effects of consumption smoothing over the life cycle. For instance tertiary students will have very low incomes for a short period of time, but their expected lifetime earnings may be quite high and thus they may borrow while their income is low to finance a higher level of current expenditure. A similar story is true for the retired population. Their current income flows may be low, but they may draw on savings that they accumulated in the past when they were on high incomes in order to sustain a higher level of expenditure when they retire. Thus, we expect the characteristics of low income and expenditure groups to differ.

In the second, third and fourth columns of figures of Table 3.1 we show the characteristics of households drawn from the bottom quintile of equivalent disposable income, the bottom quintile of equivalent total expenditure and from the population as a whole. We have normalised the effect of size and composition of the households by constructing equalised income and expenditure distributions. This is achieved by dividing the raw income and

expenditure by the square root of the number of members in the household. We have also undertaken this comparison with raw data and these are reported in Appendix 1.

Eight categories of characteristics are shown; household family composition, gender of household head, age of household head, labour force status of the head of the household, principal source of income, average household expenditure and income, presence of dependent children, and finally whether cash flow problems were apparent and the effect of these on the particular population. First we consider those with zero and negative income.

3.2 Households with zero or negative income

There are a number of quite plausible circumstances in which zero or negative household income can be explained – business loss, transition between jobs, education and/or receipt of benefits, and unpaid leave may be examples. What is important is the way in which we interpret this information and whether current non-positive income is an appropriate indicator of poverty or disadvantage. The HES data attributes implausibly high marginal and average propensities to consume to households at the bottom end implying unsustainable rates of dissaving. For instance Harding and Greenwell (2002) show that the ratio of equivalent total expenditure to equivalent disposable income for the lowest income decile over the period between 1984 and 1998/99 was between 2.0 and 2.5. Many households in this bottom decile are self-employed or have heads of retirement age. But what about the others? The HES does not record the running down of assets to finance expenditure or irregular receipts. The latter may include proceeds from the sale of a large durable item or lump-sum superannuation or compensation pay out. The ABS (1996, p.10) stresses that the difference between recorded income and expenditure is not to be regarded as a measure of savings.

The first column of Table 3.1 shows that households with zero or negative income do not exhibit the characteristics of conventionally poor households. They do appear to contain the types of households in which there is a transition such as is described above. The last block of data describing indicators of financial distress is revealing. In general these households exhibited less distress than the population as a whole and much less distress than is revealed for the bottom quintile of income or expenditure. Lower proportions went without meals, were unable to heat their home, sought financial help from friends and had cash flow problems. A negligible proportion sought assistance from welfare organisations and the only indicator of distress in which they exceeded population and low income and expenditure quintile distress was in payment of registrations and insurance. It seems likely that this group does indeed largely represent households in transitional situations. While the sample may include households of genuinely low income it is not possible to identify these. For trend analysis it is probably safest to omit them from conventional poverty and inequality measurement.

There are also households with negative expenditure, possible because expenditure includes some forms of expenditure on durables. If a household sells a car for example, then a negative expenditure item may be entered. This makes using low expenditure as an indicator of poverty problematic as most of these households are by no means in the same circumstance as the truly poor. Blacklow and Ray (2000) for instance treat these negative incomes as income flows and thus add the amount of negative expenditure to total income and set the expenditure item to zero. In the following analysis we leave those with negative expenditures in the sample, noting that this is an issue that needs to be addressed if expenditure is used to measure living standards. As capital expenditure items are not present in our expenditure measure the presence of negative items will not be as significant a problem as would be the case if these items were included. The issue should however be kept in mind.

3.3 Comparing households in the bottom equivalised income and expenditure quintiles

Table 3.1 shows that there are indeed differences in the characteristics between the bottom income and expenditure quintiles. The demographic composition of the bottom income

quintile varies from that of the bottom expenditure quintile with a higher tendency for sole parents and single-adults living alone to be represented in the low-income group. Households with a head aged 65 years plus are more likely to be in the low-income group. Female-headed households are also more likely to be on low incomes reflecting the higher proportion of sole parents and older people living alone. If expenditure is acting as a proxy for permanent income, part of these differences are expected as the older, retired, population may be drawing down on savings to keep their expenditure at levels sustaining their higher levels of lifetime consumption. Sole parents, however may be a group that expenditure estimates do not pick up, as the presence of children makes it difficult for them to lower their expenditure much even if they do have very low levels of income. This may be an example where expenditure based measure fails to capture those most in need. Over all household types, the proportion of households with dependent children in the bottom income quintile does not vary significantly with that of the lowest expenditure quintile.

We also note quite stark differences between the characteristics of the bottom quintiles and the population as a whole. Lone persons constitute nearly half of the bottom income quintile and nearly 40 percent of the bottom expenditure quintile, a much larger representation than in the population as a whole with around 24 per cent of one-person households. Couples with and without children are underrepresented in the bottom quintiles relative to the population. Sole parents are over represented in the low-income quintile. Households with heads aged over 65 make up a much larger proportion of the low income and expenditure groups relative to the overall population. Households headed by females are over represented in the low-income quintile, reflecting the fact that single person households over the age of 65 make up a large share of the bottom quintiles. The proportion of couple households with dependent children in the lowest quintiles, regardless of the measure used, is much lower than the population average.

Table 3.1 Comparison of households in the bottom equivalent income quintile, the bottom equivalent total expenditure quintile and in the sample as a whole, 1998/99¹

Characteristic	Zero or Neg. income	Bottom quintile ² Income	Expenditure	Whole sample ²
Household family composition, %				
Lone person	42.0	48.3	37.7	24.0
Couple only	26.3	21.0	21.0	24.6
Couple with dependent children	18.7	14.2	27.8	33.5
One parent with dependent children	3.3	14.2	8.5	8.5
Other	9.7	2.4	8.5	9.5
Gender of household reference head, %				
Male	46.6	38.8	62.6	61.0
Female	53.4	61.2	37.5	39.0
Age of household reference head, %				
Less than 25 years	4.2	4.6	4.2	5.4
25 to 64 years	79.6	58.0	70.1	75.5
65 years or older	16.3	37.4	25.8	19.2
Presence of dependent children (%)				
Dependents present	20.1	26.7	30.6	35.9
Labour force status (%) (Household reference head)				
Wage and salary – full-time	11.8	2.1	46.0	49.1
Wage and salary – part-time	13.0	5.9	5.2	9.4
Self-employed	34.9	6.8	6.7	7.3
Unemployed	2.7	10.3	3.6	2.8
Not in the labour force	37.7	74.9	38.5	31.5
Principal source of income (%)				
Government benefits		81.3	35.8	28.6
Wage and salary		6.7	49.8	57.5
Self-employed		3.4	7.5	6.3
Superannuation, investment, other		8.3	7.0	7.5
Income and expenditure				
Total household expenditure (\$pw)	602.8	267.0	88.2	328.6
Household disposable income (\$ pw)	-468.3	174.8	648.0	566.7
Total household income (\$pw)	-480.1	172.8	489.6	453.1
Cash flow problems (%)				
Could not pay electricity/gas/telephone bill	13.7	25.4	11.9	16.1
Could not pay registration/insurance	9.5	8.6	3.8	6.5
Went without meals	1.7	6.7	2.3	2.7
Unable to heat home	0.0	5.6	2.3	2.3
Sought assistance from welfare organisation	0.1	9.4	3.5	3.5
Sought financial help from friends	5.5	16.0	7.1	10.0
Had cash flow problems in past year	19.5	33.6	16.7	22.1

Notes: Income quintiles are based on equivalised household disposable incomes. Expenditure quintiles are based on equivalised household expenditure on goods and services minus expenditure on income tax. The equivalence scale used was the square root of household size.

In other respects, such as labour force status of the household head, and principal source of income, the low-income and expenditure quintiles vary. Households with wage and salary earning heads are much more likely to have low expenditure, but not low income. Households where the head is unemployed or not in the labour force are represented more heavily in the low-income quintile relative to the low-expenditure quintile, with those not in the labour force much more likely to have low incomes but not low expenditures. As we have omitted those

with zero or negative incomes, households with self-employed heads are not heavily represented in the lowest income/expenditure quintile relative to the general population. However, households with their principal source of income from government benefits or superannuation, investment or other sources are much more apparent in the lowest income quintile, and households with most of their income from employment, particularly from wage and salary income, more evident in the bottom expenditure quintile.

There are also large differences in average incomes and expenditures of households in the bottom income and expenditure quintiles. While the average household in the bottom disposable income quintile has equivalised disposable income of \$175 per week, equivalised expenditure is \$267 per week implying a dissaving of around \$90 per week.¹¹ By contrast the average household in the bottom expenditure quintile has disposable income of \$648 per week and expenditure of \$88 per week implying a saving of nearly \$550 per week. Such a low level of average expenditure is most certainly due to households with negative expenditure remaining in our sample. Further research is needed to look into the group with negative expenditures.

In considering welfare implications the issue is whether concern rests on those with insufficient income to meet their commitments, or on those with, apparently, the meanest standard of living. In the first circumstance the emphasis is on consideration of the opportunities available to households whereas in the second it is on the actual circumstance irrespective of opportunities or how those circumstances were brought about. From a welfare point of view, arguments could be made for concentrating on those in the first category on the grounds that if those in second category wanted to improve their situation they would need only to spend more.

Further indication of differences in standards of living between households with low incomes relative to those with low expenditures can be gained by looking at various variables indicating financial distress within households at the time of the interview. As we can see, from the bottom panel of Table 3.1, the low-income group appears to struggle much more to make ends meet. Nevertheless the low-expenditure group suffers more than the general population in terms of the more extreme forms of financial distress such as going without meals, not being able to heat their home, and seeking assistance from welfare organisations. However, a larger proportion of those on low incomes are adversely affected, with many more of these households also finding it difficult to pay their bills and seeking financial help from friends. Overall, low-income households were more likely to have had cash flow problems in the past twelve months than their low expenditure counterparts. The low-income households seem to be much closer in characteristics to the image of a target group of welfare concern described in section 2.1.

3.4 Households with low income but not low expenditure

To gain a deeper understanding of the differences between low-income households relative to low-expenditure households we now turn to examining the characteristics of households in the low-income group that are not in the low-expenditure groups and vice versa. Table 3.2 reports data on households that are low income but not low expenditure and households that are low expenditure but not low income. Again we have equivalised the raw data so that we may better compare across households and draw some inferences with reference to our views about neediness discussed in relation to groups in section 2.1.

According to the life-cycle/permanent income hypothesis, those with low incomes and high expenditures are experiencing a temporary fluctuation in their income, with a level of permanent income high enough to sustain their higher level of expenditure. The mirror image

¹¹ Johnson et al (1998) note that it is difficult to believe that such high weekly propensities to consume with such high levels of dissaving for this income group is sustainable. As we have noted the difference may occur for a number of reasons; households may be drawing on accumulated assets, they may be supported by transfers from other households, or there may be undeclared income.

is said to occur in households with high incomes and low expenditures, their current state of high income is foreseen as a temporary one and thus with consumption being smoothed over their lifetime, their level of expenditure in the current period more closely relates to their lower level of expected permanent income. The former case would largely consist of households where the head is temporarily out of work, a student and borrowing in the expectation of higher returns on their education in the future, or retired and dissaving to sustain their higher levels of expenditure. The latter case may include those in short-term or temporary employment.

Table 3.2 Characteristics of lowest income but not expenditure quintiles and vice versa^{1 2}

Characteristic	Lowest income but not expenditure quintile	Lowest expenditure but not income quintile	Low income and low expenditure quintile
Expenditure/Income quintile			
Second	41.5	20.0	-
Third	26.0	19.6	-
Fourth and fifth	32.5	60.5	-
Household type			
Person living alone	42.7	27.6	61.7
Couple only	23.2	23.3	15.6
Couple with children	16.7	36.0	8.2
Lone parent	14.9	6.8	12.4
Other	2.5	6.3	2.2
Other demographic			
Female	61.0	27.4	61.5
Dependents under 25	29.5	35.0	20.1
Retired persons in hh	37.1	18.0	49.7
Age of head			
Under 25 years	4.7	4.1	4.4
25 to 65 years	62.1	79.2	48.4
65 years plus	33.3	16.8	47.2
Marital status of head			
Never married	16.3	17.3	20.8
Widowed/divorced/separated	40.8	20.5	53.2
Married/defacto	43.0	62.2	26.0
Principal source of income			
Wages and Salaries	8.5	69.5	2.5
Self employed	4.1	9.9	1.7
Other private income	9.5	7.7	5.3
Benefits	77.5	12.9	90.5
Labour force status of head			
Full-time employee	2.6	64.9	0.8
Part-time employee	7.0	6.1	3.3
Self employed	8.6	8.5	2.3
Unemployed	10.3	0.8	10.3
Not in labour force	71.5	19.8	83.3
Average weekly income/expenditure			
Expenditure	325.7	72.1	126.6
Disposable Income	170.3	619.7	179.0
Financial distress			
Could not pay gas/electricity/phone bill	27.1	7.9	21.5
Could not pay registration/insurance	9.5	2.7	6.6
Went without meals	7.6	1.3	4.6
Unable to heat home	5.7	1.2	5.1
Sought assistance	9.3	1.0	9.7
Sought financial help from friends	17.3	4.6	13.1
Had cash flow problems in past year ³	64.6	11.4	29.3
Number of households in sample	920	939	379

Notes: (i) Income quintiles are based on equivalised household disposable incomes. Expenditure quintiles are based on equivalised household expenditure on goods and services minus expenditure on income tax. The equivalence scale used was the square root of household size. (ii) Households with zero and negative incomes have been omitted. (iii) "Had cash flow problems in past year" is a composite variable derived from the outlined financial distress variables.

So what does Table 3.2 tell us? Do the characteristics of those with low incomes and higher levels of expenditure and those with high incomes and low levels of expenditure support the theory? Does it seem feasible that expenditure can be used to measure those most vulnerable in society over the long term?

The first block of characteristics of the group in the low income but not the low expenditure quintile shows that over forty percent were in the very next quintile, the second, on the basis of their expenditure. Undoubtedly a fair proportion of these may have had expenditure only a little bit higher than income and their definition derives from the arbitrariness involved in the selection of the threshold dividing the first and second quintile. However, more than half were more than one quintile removed indicating that for these the classification is not a mere artefact of group definition. This is confirmed by considering the average weekly income and expenditure of this group. Average expenditure was more than double income. As was stated earlier, plausible explanations include that many of these households were in some form of transient income situation – where perhaps the income earners in the household were between jobs, were suffering some temporary downturn in a business, were between education and work, were between work and retirement and so on. A second plausible explanation is that they are households in which savings (and other assets) are being used up.

The proportion of self-employed was slightly higher than for the low income/low expenditure population as a whole suggesting that some temporary situation in business could explain only a small proportion of the apparent large dissaving. Of more interest was the very high proportion of retired people in the sample. The sample was in general quite old, with over double the proportion of households with heads over 65 as the population at large (shown in Table 3.1), nearly double the proportion of single people and overwhelmingly headed by a female. Overwhelmingly the group was not in the labour force and dependent on government payments.

The levels of financial distress were higher than from the population with low incomes and low expenditure. Distress was also higher. Nearly 27 percent had difficulty with the utilities bill compared to a population average of 21 percent for households with both low income and low expenditure and 16 percent for the population as a whole. Nearly two-thirds had had some form of cash flow problems over the year compared to a fifth in the whole population. However the level of debt was lower than for the population as a whole suggesting that the higher expenditure was not being funded by debt but rather by previous savings.

The issue is, are the results believable for this group? Is it plausible for such a group to spend each week, double their stated income? Undoubtedly there will be some individuals for whom this is true. However the real issue in interpreting the results is whether the magnitudes are believable. The alternative explanation is that the results are generated partly by under-reporting of income. That in fact the income stated is not the true income. At the time of the survey (December 1998) the old-age pension for a single female was \$178.65 per week (Melbourne Institute, 1999). If the pensioner were living in private accommodation they would be eligible for an additional \$37.60 in rent assistance. We cannot compare these amounts with the figures in Table 3.2 because they are equivalised but the raw figures are shown in Appendix 1. The stated average income for the group as a whole (bearing in mind that it would include non-pensioners who would, in general, have higher income) was \$215 per week. The stated income seems plausible. It is the average expenditure that seems higher than it should for this group not the income. Given the method of gathering information about expenditure (the HES used a diary for gathering information about consumables, plus recall for durable items and for other major items of expenditure such as cars and houses) the scope for large over-estimation of expenditure seems unlikely. However plausible imbalance

between income and spending would occur where assets are being run down (for instance old people drawing on savings to maintain a higher standard of living than their income would allow). This explanation would most fit the data and in this case there are no welfare implications.

In summary while the group undoubtedly includes some households who were in poverty, in the main it probably represents older people who were legitimately running down their assets in order to maintain a higher standard of living.

3.5 Households with low expenditure but not low incomes

Now turning to those with low levels of expenditure but not on low incomes. The second column of Table 3.2 shows us that the divergence with the quintile of income for low expenditure households is much greater than in the case of the bottom income quintile. Here, only around a fifth of low expenditure households have incomes in the next quintile, with over sixty percent with incomes in the fourth and fifth quintiles. More couples with children, and less persons living alone and lone-parent households make up this group than in the low income/higher expenditure case suggesting perhaps that higher income families with children may be more risk averse in terms of adverse income shocks and thus save much more for their families future. Females head less of these low expenditure/higher income households, and less have elderly, retired heads. Reflecting the lower proportion of the retired, a larger portion of the low expenditure/higher income households have either never been married or are married/defacto. Also, unlike the low income/higher expenditure households largely dependent on government benefits and other unearned private income, a full-time employed member receiving wages and salaries largely heads these low expenditure/higher income households. Households with unemployed heads are virtually non-existent in this category and households with other jobless heads are significantly less represented in this category. While average expenditures were incredibly low, average equivalised net income was over \$620 a week. Not surprisingly, this group is not subject to anywhere near as much financial distress as the low-income households, nor as the population as a whole. While the pattern of expenditure and income may be puzzling it is of little welfare significance however, we may still expect some of these households to be in some financial distress if their high levels of income are transitory.

3.6 Households with both low income and low expenditure

Single, elderly retired females living alone seem to make up the majority of households with both low incomes and low expenditure. This is generally a feature of many of the low-income households, but not of the low expenditure households, which may be explained by the fact that households with negative expenditure remain in our sample of low expenditure households. Also, the proportion of households with dependent children consistently represented in the lowest quintiles is much lower than the other population groups. This group appear to live within their means; expenditure is comfortably less than income on average.

3.7 Summary

In sum, there are three key findings from this section. The first is that households with zero or negative income do not exhibit the characteristics of conventionally poor households. It seems likely that this group does indeed largely represent households in transitional situations and should be left out of conventional poverty measurement. Second, the category of households that were in lowest quintile on the basis of their income but not in the lowest quintile on the basis of their expenditure seems primarily to be older people who were legitimately running down their assets in order to maintain a higher standard of living. Some households in this category are undoubtedly experiencing poverty but this a small fraction of the total. Third, that category of households that fall into the lowest quintile on the basis of expenditure but not on the basis of income appears to be mainly a group with an above average propensity to save. Their tendency to spend at below average rates does not appear to be driven by financial distress in the manner expected of a low income, impoverished

household. This may mainly be due to households with negative expenditure remaining in our population of low-expenditure households, thus further research needs to be undertaken to evaluate the effect of their removal from the sample. The overall picture is of a heterogeneous population in the low income and expenditure quintiles many of which do not exhibit the characteristics of the conventionally needy. Accordingly we should be wary about basing welfare policy on the apparent income and/or expenditure situation of these households.

4. Government transfers in administrative data and income surveys

While in the previous section we examined the differences between income and expenditure in the HES we now turn to the ABS income surveys and try to determine whether non-reporting or under-reporting of income is apparent. We have turned our focus to these surveys because they provide a much more detailed disaggregation of income and are seen to provide a more accurate measure of income than the HES. The ideal scenario would be to have data on the true level of income that each person in the population receives against which we could compare our sample estimates. But this is not possible given the limitations of existing statistical knowledge. However we do have the closest thing possible to true incomes for one subgroup of the population: the recipients of FaCS income support payments. So in this section we compare the ABS data on observed benefit recipients with FaCS administrative data records to determine the discrepancies between numbers and types of recipients and the differences in the distribution of incomes for the observed benefit population. Directly comparing the Department's administrative records of actual benefit recipient numbers and amounts with the respondent numbers reported in the Australian Income and Housing Costs Survey (SIHC) provides insight into the degree of error in the SIHC survey.

The following analysis examines data from the benefit payment records of the Department of Family and Community Services from January 1995 to June 2000 (the LDS) alongside data from the SIHCs conducted from 1994/95 to 1997/98.

4.1 Detailed explanations of the data sources

The Department's administrative data records we use are based on a 1% sample of all individuals in receipt of a FaCS income support payment at any time throughout the period between January 1995 and June 2000. DVA payments are not included and Austudy/Abstudy payments prior to July 1998 are not included. Income support payments also exclude basic Parenting Payment and Minimum Family Payment. An observation is available for each fortnight that individuals are observed on an income support payment. Data on the partners of recipients of income support payments are also part of the data set, with identifiers enabling the matching of individuals to their partners in order to construct measures of their combined incomes.

The Australian Survey of Income and Housing Costs (SIHC) has been conducted since 1994/95 as part of the Monthly Population Survey and contains detailed unit record data on the composition of income and housing costs both at the income unit level and at the person level. Around 650 households are surveyed monthly. Prior to 1994 similar information was collected as part of the Income Distribution Survey (IDS), which was conducted at four yearly intervals over 3 months in the second half of the year. At present there are unit record data for the following years: 1982, 1986, 1990, 1994/95, 1995/96, 1996/97 and 1997/98.

Demographic characteristics of each person of workforce age in each income unit are recorded, including: age, sex, marital status, country of birth, number of dependent children and age of youngest child. Other characteristics recorded include: dwelling type and structure, tenure type, current weekly rent paid and current weekly loan repayments of each income unit, employment status, labour force status, highest educational qualification, weekly hours of work, occupation and industry in main job, duration of unemployment, current weekly earned and unearned income from various sources and annual income from each source in previous financial year. Income sources are also detailed and include income from wages and

salary, property and interest, social security allowances and pensions, superannuation and other regular sources.

A caveat is that our data source only contains information on income support payments. Thus individuals or families on the basic parenting payment or the minimum level of family payment are not captured here. The benefits that are included in this study include Newstart Allowance (NSA), Job Search Allowance (JSA), Mature Age Allowance (MAA), Youth Training Allowance, Age Pension, Carer Pension, Wife Pension, Disability Support Pension (DSP), Sole Parent Pension, Sickness Allowance (SA), Widow Allowance, Partner Allowance and Special Benefit.

4.2 Recipient numbers

Comparisons of recipient numbers by payment types are presented in Tables 4.1 and 4.2. Table 4.1 provides point in time population estimates of recipient numbers on an income support payment in 1997/98. The estimates across the years between 1994/95 and 1996/97 are presented in Appendix 1. In the SIHC data we rely on a person's usual current weekly income from each respective benefit type while in the LDS each person must be registered on a particular benefit and in receipt of benefit income from this source. In the LDS our estimate (which from this point on we refer to as the *actual* or *true* value) is constructed by taking the average number of recipients by payment type over the year. The first two columns in Table 4.1 present the weighted recipient numbers by benefit type from the SIHC and LDS respectively. The absolute difference in recipient numbers by payment type between the two data sets are presented in the third column, i.e. column 1 minus column 2, while the final column presents the difference as a ratio i.e. column 1 divided by column 2.

Table 4.1: Weekly/fortnightly recipient numbers by payment type, 1997/98

	SIHC	LDS	Difference	Ratio
Total	3,434,615	3,649,246	-214,631	0.94
NSA/JSA	582,930	722,158	-139,228	0.81
Age Pension	1,653,201	1,671,723	-18,522	0.99
MAA	53,839	54,269	-430	0.99
DSP	487,498	536,488	-48,990	0.91
Sole parent pension	326,276	352,096	-25,820	0.93
Wife/carer pension	130,755	150,673	-19,918	0.87
Sickness allowance	48,603	12,323	36,280	3.94
Widow Allowance	43,324	39,031	4,293	1.11
Special benefit	12,788	9,908	2,880	1.29
Partner allowance	73,971	71,423	2,548	1.04
Youth training allowance	27,749	29,154	-1,405	0.95

Over the four years we see that the SIHC does not perform too badly in terms of identifying the selected benefit recipients based on a point in time, with between 90 to 94 per cent of actual recipients identified. Investigating further by payment type we see that recipients of unemployment benefits (Newstart Allowees combined with Job Search Allowees in earlier years and Youth Training Allowees which we do not concentrate on individually as they are such a small sample in the ABS survey) are significantly understated in the ABS surveys with only between 81 and 90 per cent of recipients of NSA/JSA identified. Sole-parent pensioners are also understated, not quite to the same extent in the last two years of the survey, but it is evident that there are problems with estimating the sole-parent pensioner population at a particular point in time in the ABS survey. The age pensioner population seems fairly accurate, particularly if one considers that the SIHC does not include those in nursing homes. Disability support pensioners are consistently understated, however sickness allowees are grossly overstated and it seems likely that many of those surveyed have mistakenly stated that

they were in receipt of a sickness payment rather than the disability support pension. In fact, the SIHC's estimate of the population in receipt of either DSP or SA is fairly accurate. The remaining payments have very small sample populations in the SIHC, which explains the volatility in the ratios presented over the years. These ratios are roughly consistent with studies performed internally by the ABS (see ABS, 2000b).

While population estimates of the selected benefit recipients based on current weekly income are slightly understated over the years from 1994/95 to 1997/98, when one examines the estimates based on income from the previous financial year, the differences between administrative data sources and the ABS survey are much more pronounced. This is consistent with the findings of Tseng and Wilkins (2002). In Table 4.2 (below) we present the population estimates of recipients of the benefits outlined earlier over the financial year 1996/7 (Table A4 in Appendix 2 presents the estimates for the 1995/96 financial year). The estimates are calculated from the SIHC using the provided information on annual incomes from the previous financial year and then weighting the data using the population weights provided. Thus, in the 1997/98 survey respondents were asked to provide the details of their income from the 1996/7 financial year. In the LDS data, since we have fortnightly payment details for each individual, we know if a person has been in receipt of a particular payment at any point throughout the year. As we have a 1 per cent sample, we then multiply the sample estimate by one hundred to obtain our estimate of *actual* recipient numbers.

Current information from respondents (at the time of the interview) enables identification of over ninety per cent of the selected benefit population, Table 4.2 shows that using retrospective (annual) income information reduces this percentage by quite a significant amount. In fact, using retrospective annual income information, we can identify less than eighty per cent of recipients of the set of benefits with which we are concerned.

Table 4.2: Annual recipient numbers by payment type, 1996/97

	SIHC	LDS	Difference	Ratio
All benefits	3,607,820	4,563,500	-955,680	0.79
NSA/JSA	910,730	1,450,300	-539,570	0.63
Age Pension	1,492,662	1,767,400	-274,738	0.84
MAA	69,087	75,100	-6,013	0.92
DSP	481,006	566,700	-85,694	0.85
Sole parent pension	344,907	441,700	-96,793	0.78
Wife/carer pension	130,697	182,800	-52,103	0.71
Sickness allowance	69,418	67,300	2,118	1.03
Widow Allowance	54,498	76,300	-21,802	0.71
Special benefit	18,316	37,800	-19,484	0.48
Partner allowance	81,711	102,900	-21,189	0.79
Youth training allowance	25,290	84,500	-59,210	0.30

Here, all of the major benefits are much more underestimated than was apparent in Table 4.1. It is also evident that it is the benefit types that tend to have shorter durations that suffer most, a finding consistent with other studies such as Tseng and Wilkins, forthcoming. In particular, only around sixty per cent of recipients of Newstart/Jobsearch Allowance are represented in the SIHC. As it may be close to two years between a person's receipt of a payment and the survey interview, it seems likely that the cause of much of the discrepancy is due to recall error.

For instance, say a person is interviewed in June of 1998 and thus will be entered as participating in the 1997/98 SIHC. Suppose that this person was on Newstart Allowance for only a month or two over June/July of 1996. This person will surely not be able to recall the

amount of the benefit received over the 1996/7 financial year, in fact they will most likely not even mention receiving the benefit at all. The time lag contributes to the measurement error.

What the information in Tables 4.1 and 4.2 highlights is that while income may be understated in the SIHC due to the under-representation of the benefit population, *current* income provides a much better estimate of the benefit population than annual income.

So far we have found that the SIHC data under-estimates the benefit population, with particular benefit types more at risk of being under-estimated than others. In Table 4.3 we take the benefit populations generated by the information on current weekly/fortnightly details, which were shown in Table 4.1, and look at the characteristics of these estimated populations across all benefits. In Table 4.4 we concentrate on the major payment groups to determine whether there are any groups that are consistently being under-represented which may explain the discrepancy in the aggregate recipient numbers. We only present the results for 1997/98, as there are no significant variations over the years.

In Table 4.3 we first present the characteristics of the general benefit population with the results from the SIHC in the first column, and the contributions of the demographic groups in the LDS presented in the second column. From the characteristics of the general benefit population, the result that is most apparent concerns marital status. Benefit recipients that are not married or in a defacto relationship are quite significantly *under*-represented in the ABS survey. This is seen consistently across all survey years. Males are slightly under-represented, although it is unlikely that this difference is statistically significant.

Table 4.3: Characteristics of benefit recipients, all benefits, 1997/98

	SIHC	LDS
Male	42.5	43.1
Female	57.5	56.9
Under 25 yrs	7.7	9.2
25 to 34 yrs	10.5	10.9
35 to 44 yrs	10.2	10.4
45 to 54 yrs	10.3	10.8
55 to 64 yrs	18.6	17.9
65 yrs plus	42.7	40.8
Married	49.8	42.5
Australian born	67.0	68.6
Overseas born	33.0	21.3
Children under 15	15.6	15.1
Age of youngest child		
0 to 5 yrs	8.4	7.2
6 to 14 yrs	7.2	7.8
Total	3,434,615	3,649,246

The young seem under-represented, with the main difference in the group under 25 years. Perhaps this is due to mobility: young singles are more likely to move around and the ABS may find it difficult to locate these people. The young, unmarried benefit population is also more likely to be on benefit for a shorter duration and thus may be less likely to report their benefit receipt.

The picture clears up a little when we delve deeper into the benefit population and examine the characteristics of benefit recipients by individual payment types. Table 4.4a and 4.4b

presents the characteristics of the recipients of some of the more major social security client groups, i.e. NSA/ JSA, Age Pension, DSP in Table 4.4a and Sole Parent Pension, Sickness Allowance and Youth Training Allowance in Table 4.4b.

Sample numbers for recipients of Youth Training Allowance are low and thus we should not place too much weight on the results in regard to this particular payment. Characteristics of recipients of the payments presented in Tables 4.4a and b generally follow similar patterns to the general case. However the payment that seems to exhibit the most pronounced differences is the NSA/ JSA (which includes the Jobsearch Allowance), particularly with relation to marital status.

Table 4.4a: Characteristics of major income support payment types, 1997/98

	NSA/ JSA		Age Pension		DSP	
	SIHC	LDS	SIHC	LDS	SIHC	LDS
Male	69.5	68.4	36.7	36.0	62.5	66.6
Female	30.5	31.6	63.3	64.0	37.5	33.4
Under 25 yrs	27.9	30.2	-	-	4.5	5.6
25 to 34 yrs	26.3	27.5	-	-	10.7	10.2
35 to 44 yrs	21.4	19.2	-	-	15.4	16.2
45 to 54 yrs	13.6	15.8	-	-	28.8	27.5
55 to 64 yrs	10.8	7.3	12.5	11.8	40.4	40.0
65 yrs plus	-	-	87.5	88.2	0.2	0.5
Married	45.2	27.9	58.4	51.9	46.0	41.2
Australian born	67.4	75.5	63.8	64.3	69.4	69.9
Overseas born	32.6	24.3	36.2	35.7	37.5	30.0
Kids under 15	23.5	15.3	0.1	0.5	9.5	9.3
Age of youngest child						
0 to 5 yrs	16.2	8.9	-	0.1	3.3	3.0
6 to 14 yrs	7.3	6.4	0.1	0.4	6.2	6.3
Total	582,930	722,158	1,653,201	1,671,723	487,498	536,488

Table 4.4b: Characteristics of major income support payment types, 1997/98

	Sole parent pension		Sickness Allowance		Youth Training Allowance	
	SIHC	LDS	SIHC	LDS	SIHC	LDS
Male	6.3	6.9	80.5	64.7	48.1	51.0
Female	93.7	93.1	19.5	35.3	51.9	49.0
Under 25 yrs	15.0	15.6	7.4	17.6	100.0	100.0
25 to 34 yrs	41.8	38.4	26.7	22.5	-	-
35 to 44 yrs	32.6	35.7	21.6	21.1	-	-
45 to 54 yrs	10.6	9.6	32.3	25.8	-	-

Understanding and improving data quality relating to low income households

55 to 64 yrs	0.1	0.7	11.9	13.0	-	-
65 yrs plus	-	0.1	-	-	-	-
Married	1.3	0.3	39.8	35.6	3.6	6.0
Australian born	80.6	79.9	75.1	78.0	93.3	93.4
Overseas born	19.4	20.1	24.9	22.0	6.7	6.7
Kids under 15	96.6	99.9	23.1	15.6	-	0.5
Age of youngest child						
0 to 5 yrs	50.5	49.5	13.9	10.1	-	0.4
6 to 14 yrs	46.1	50.4	9.1	5.5	-	0.1
Total	326,276	352,096	48,603	12,323	27,749	29,154

Single Newstart Allowees seem to be significantly under-represented in the SIHC. There are a number of explanations for this. One is that recipients of Partner Allowance may be reporting receipt of Newstart Allowance if their partner is on this payment, thus bolstering the share of the married population in the survey data. However this should have the effect of increasing the entire Newstart population, and we know from the analysis above that the Newstart population has been consistently under-represented in the SIHC data. That is not to say that recipients of Partner Allowance are not reporting they are on Newstart Allowance - they may be. But, if they are, then the population of Newstart recipients is significantly more under-represented than we originally thought.

Another explanation is mobility of single people, discussed above. A third explanation is the shorter average duration of Newstart recipients. Newstart recipients are more likely to be receiving benefits for shorter terms than for those of other major income support payments, particularly those receiving pensions. This shorter duration may mean that respondents to the SIHC are more likely to fail to mention their benefits - it simply is not as strongly ingrained in their personal picture of their own income stream.

4.3 Total income from benefits

We now turn to examining incomes across the observed benefit population. This contrasts with the analysis above where we looked at the *difference* in recipient numbers between the two data sources. In this section we investigate *total benefit* incomes with the purpose of determining whether there is any evidence of under-reporting of incomes for the observed benefit population. This section uses the individual as the unit of analysis. The SIHC includes measures of current weekly benefit income by benefit type and are thus doubled to compare with the fortnightly income information available in the LDS.

Comparisons of total benefit income by payment types are presented in Tables 4.5 and 4.6. Table 4.5 provides point in time population estimates of total benefit incomes in 1997/98. The estimates across the years between 1994/95 and 1996/97 are presented in Appendix 3. The first two columns in Table 4.5 present weighted estimates of total benefit income by payment type from the SIHC and LDS respectively. The absolute difference in recipient income by payment type between the two data sets are presented in the third column, i.e. column 1 minus column 2, while the final column presents the difference as a ratio i.e. column 1 divided by column 2.

As was seen above, the SIHC identifies between 90 to 94 per cent of benefit recipients at any given point in time. If one turns to the total amount of income reported to be received from benefits, understating of income seems to be apparent with only around 82 to 85 per cent of total income from benefits recorded. When breaking this down by payment type, the same payments vulnerable to being underrepresented in terms of population estimates are also most prone to understating of benefit income. Total income from Newstart and Job Search Allowance is considerably understated in the ABS surveys with only 73 to 86 per cent of this income reported. The earlier surveys in 1994/95 and 1995/96 were more successful in

identifying income from NSA and JSA with the 1996/97 and 1997/98 surveys only identifying 73 per cent of NSA/JSA income. Income from the sole-parent pension is also understated, with the proportion of income identified as low as 66 per cent and as high as 79 per cent. Although the age pensioner population seems to be represented quite accurately in terms of its aggregate population in the ABS surveys, income from Age Pensions is considerably understated with 86 to 89 per cent of income from Age Pensions observed in the SIHCs. As the recipient numbers show there seems to be some confusion as to whether certain individuals are in receipt of Disability Support Pension or Sickness Allowance. However, even when combining these benefits the SIHC understates incomes from DSP and SA. The remaining payments have very small sample populations in the SIHC, which explains the volatility in the ratios presented over the years. Again, these ratios are consistent with internal ABS studies.

Table 4.5: Total fortnightly benefit income by payment type, 1997/98, \$'000s

	SIHC	LDS	Difference	Ratio
Total	985,490	1,159,485	-173,995	0.85
NSA/JSA	162,779	221,771	-58,992	0.73
Age Pension	462,360	518,352	-55,992	0.89
MAA	15,005	16,749	-1,744	0.90
DSP	152,749	183,971	-31,223	0.83
Sole parent pension	101,677	129,166	-27,489	0.79
Wife/carer pension	35,098	44,169	-9,071	0.79
Sickness allowance	14,835	3,838	10,997	3.87
Widow Allowance	13,415	13,202	213	1.02
Special benefit	3,375	3,188	187	1.06
Partner allowance	19,490	20,280	-790	0.96
Youth training allowance	4,710	4,799	-89	0.98

Following the pattern observed with recipient numbers, using estimates of benefit income based on income from the previous financial year exacerbates the differences between administrative data sources and the ABS surveys. Estimates of total benefit income over the financial year 1996/7 by payment type are presented in Table 4.4 (below). Table A6 in Appendix 3 presents the estimates for the 1995/96 financial years.

Table 4.6: Total annual benefit income by payment type, 1996/97, \$'000s

	SIHC	LDS	Difference	Ratio
All benefits	23,486,387	28,903,428	5,417,041	0.81
NSA/JSA	4,197,100	5,833,292	1,636,192	0.72
Age Pension	10,616,232	12,680,179	2,063,947	0.84
MAA	458,742	450,178	-8,564	1.02
DSP	3,753,230	4,331,755	578,525	0.87
Sole parent pension	2,288,315	3,107,157	818,842	0.74
Wife/carer pension	874,031	1,161,099	287,068	0.75
Sickness allowance	300,993	156,339	-144,654	1.93
Widow Allowance	394,138	444,059	49,921	0.89
Special benefit	91,319	123,982	32,662	0.74
Partner allowance	463,879	491,459	27,580	0.94
Youth training allowance	48,408	123,928	75,520	0.39

As was the case with recipient numbers, using retrospective information about incomes causes benefit incomes to be understated by a larger extent than when using current income

information. What is interesting however is that the difference between using current and retrospective income information is much smaller when examining total amounts of income received from benefits. In fact, when using retrospective income, the SIHC seems to perform better at identifying total benefit income than in identifying recipient numbers. This adds weight to the view that it is individuals in receipt of small amounts of payment over the financial year that are not being picked up in the benefit population.

The information presented here combined with that of the previous subsection emphasises that benefit income is considerably understated in the ABS surveys, even if one adjusts for the under-representation of the benefit population. As was the case in identifying benefit recipients, *current* income provides a better estimate of total income from benefits than annual income.

4.4 Average incomes

In this section we investigate *average* incomes with the purpose of determining whether the average person observed to be receiving a benefit in the ABS survey represents what we expect the average persons income to be from administrative data sources. This will allow us to get a feel for whether there is any evidence of under-reporting of incomes for the *observed* benefit population. Again this section uses the individual as the unit of analysis.

Table 4.7 presents the average levels of annual benefit income observed in the SIHC and the LDS by payment type for the 1996/97 financial year. Standard deviations are also shown. Estimates for other available years are presented in Tables A7 to A9 in Appendix 4.

Table 4.7: Average annual benefit incomes by payment type, 1996/97

	SIHC		LDS	
	Mean	Standard deviation	Mean	Standard deviation
All benefits	6,510	2,712	6,334	3,113
NSA/JSA	4,609	2,852	4,022	2,853
Age Pension	7,112	2,243	7,174	2,594
MAA	6,640	2,155	5,994	2,746
DSP	7,803	2,140	7,644	2,719
Sole parent pension	6,635	2,895	7,035	3,361
Wife/carer pension	6,687	2,233	6,352	2,347
Sickness allowance	4,336	2,972	2,323	2,285
Widow Allowance	7,232	2,380	5,820	3,199
Special benefit	4,986	3,317	3,280	3,011
Partner allowance	5,677	2,294	4,776	2,583
Youth training allowance	1,914	1,487	1,467	1,430

The SIHC includes a measure of current weekly income and is thus doubled to compare with the fortnightly income available in the LDS. However the way that income is measured varies between the two data sources. In the SIHC, wage and salary income refers to usual income at the time of the interview and thus excludes any lump sum bonuses or over-time pay etc whereas the LDS includes all wage and salary income taken at the reference time. As well, in the SIHC certain benefit types are adjusted for lump-sum payments. No such adjustment occurs in the LDS. In light of these differences, we have focused on annual income because we feel that the income measures are more comparable across surveys. The difficulty with concentrating on annual income is that, as we do not observe people who have moved off benefits in the LDS, we cannot compare total incomes across the two sources.

Table 4.7 shows that on average across all of the benefits shown, the ABS SIHC reports a greater average annual income from benefits than the LDS. This is consistent with our theory that individuals in receipt of benefits for short durations have problems recalling their time on

benefit and thus not reporting their benefit income. Since these individuals receive lower levels of benefits over the year, incomes of the reported benefit population are biased upwards increasing the average.

Looking in finer detail at individual payments we can see that payments with typically shorter average durations are those that seem the most susceptible to generating higher average incomes. Payments such as NSA/ JSA, Partner Allowance, Youth Training Allowance and Sickness Allowance all exhibit a much larger average annual benefit income level in the SIHC compared to the LDS. MAA and DSP also show a large amount of variation between the two data sources. Perhaps this is due to individuals moving onto Age Pension throughout the year. As those changing benefits are more likely to be in receipt of MAA for a shorter duration than others, not reporting this income for MAA will have the effect of raising the average income as the observed MAA population will mainly consist of longer-term recipients of MAA in receipt of larger annual incomes from the benefit.

Average fortnightly benefit incomes for 1997/98 are outlined in Table 4.8. Table 4.8 indicates a consistent picture of lower average income in the SIHC compared to the LDS across payment types. This may be due to the variation across the two surveys in the way that lump-sum advances are treated. Overall, however, there is a strong indication that there is significant under-reporting of benefit incomes. The extent of this under-reporting depends on how significant lump-sum advances are to the average benefit income in LDS. Under reporting is most prevalent among sole-parent pensioners. Another category of benefit recipient with a striking result is sickness allowees whose average income was significantly greater in the SIHC in 1997/98. This may be due to confusion among recipients between receipt of DSP or SA - the DSP has a higher payment rate so this could well be an important factor driving the average up. Also, it may be due to small sample sizes as the average is quite volatile over the years.

Table 4.8: Average fortnightly benefit incomes by payment type, 1997/98

	SIHC		LDS	
	Mean	Standard deviation	Mean	Standard deviation
1997/98				
All benefits	286.93	81.57	305.51	83.67
NSA/JSA	279.24	74.11	295.28	83.74
Age Pension	279.68	84.54	298.14	80.71
MAA	278.71	70.97	296.76	64.31
DSP	313.33	70.60	329.73	73.78
Sole parent pension	311.63	86.28	352.74	90.36
Wife/carer pension	268.42	82.78	281.87	63.31
Sickness allowance	305.23	52.06	299.48	103.15
Widow Allowance	309.64	83.20	325.22	75.43
Special benefit	263.92	95.34	309.42	108.79
Partner allowance	263.49	62.42	273.02	63.07
Youth training allowance	169.73	65.03	158.28	68.25

In Table 4.9 we present average total fortnightly incomes for 1997/98 (which includes earned income, unearned income and benefit income). Table 4.9 reports average total incomes in the SIHC and LDS for each of the benefit types. Considering the results over all benefit recipients we note a closer concordance between the two data sources. Evidently under-reporting of benefit income has been made up with over reporting of non-benefit income. Alternatively there may be under-reporting of non-benefit income in the LDS. The size of benefits is dependent on non-benefit income and individuals have an incentive to under-report non-

Understanding and improving data quality relating to low income households

benefit income. While there are penalties for under-reporting it is, nevertheless likely that there will be some.

Table 4.9: Average fortnightly total incomes by payment type, 1997/98

	SIHC		LDS	
	Mean	Standard deviation	Mean	Standard deviation
All benefits	405.00	232.51	405.40	197.74
NSA/JSA	371.87	337.18	340.60	283.45
Age Pension	380.96	165.70	397.14	128.98
MAA	353.92	156.03	363.85	96.02
DSP	375.14	156.23	379.29	119.23
Sole parent pension	704.80	233.97	671.83	176.20
Wife/carer pension	376.64	182.95	392.78	181.64
Sickness allowance	367.85	174.72	321.42	121.34
Widow Allowance	382.74	95.49	395.26	91.50
Special benefit	395.32	321.66	330.70	131.82
Partner allowance	290.09	119.59	319.93	122.16
Youth training allowance	169.73	65.03	172.99	85.54

4.5 Income distributions of observed benefit population

We know that the benefit population is slightly under-represented in the SIHC, with certain payments such as NSA/ JSA and Sole Parent Pension more affected than others. We also know that total benefit incomes are significantly under-stated in the SIHC, more so than an under-representation of the benefit population implies. This suggests that an under-reporting of benefit incomes is apparent.

Now we turn to the *distribution* of incomes reported over the observed benefit population to try to determine whether there is any consistent evidence of under-reporting of their benefit income, and/or their total income. We do this by comparing the distribution of annual and fortnightly benefit incomes reported in the SIHC with those entered in the LDS over the benefit populations. In addition, we examine the total fortnightly incomes received in the two data sources. Note that we do not observe total annual incomes in the LDS because we do not have information about people once they exit the social security system.

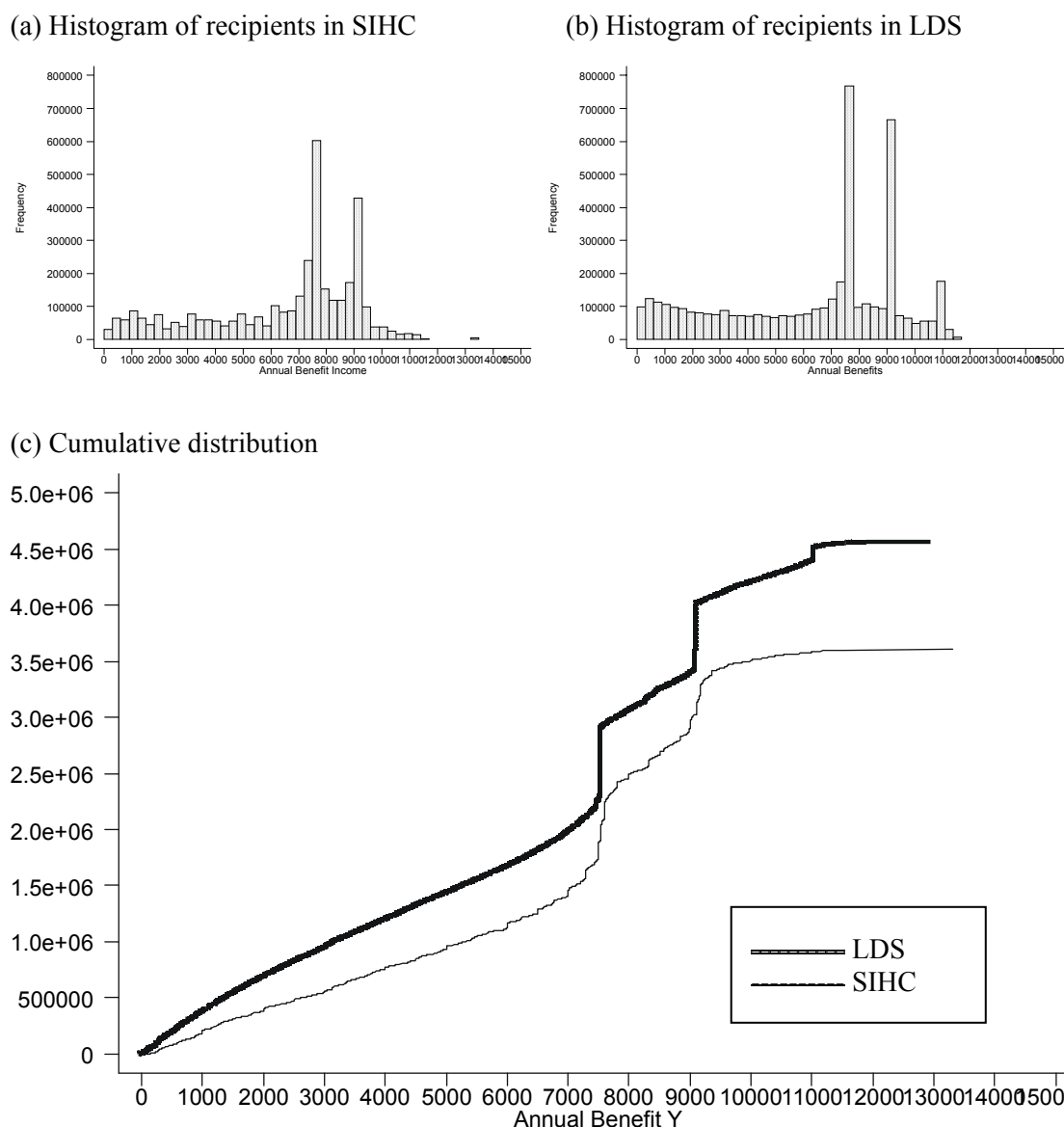
Figure 4.1 presents information about the distribution of annual income from benefits derived from the two data sources. The frequency distributions of benefit income for the observed benefit population in the SIHC and LDS are presented in Figures 4.1(a) and (b) respectively. Figure 4.1(c) presents the cumulative frequency distributions of annual benefit income across the selected benefit population for 1996/97. In Figure 4.1(c) the bold line represents the distribution of the LDS sample while the other refers to the SIHC weighted population estimates. The figure shows the weighted frequency of individuals with annual benefit incomes below the corresponding level on the x-axis. The figure shows how important it is to look at the frequency distribution of income rather than a measure of central tendency based on the observed benefit population such as the mean, or median. In Figure 4.1(c) we can see that comparing median incomes across the two data sources (representing the middle person in each distribution) would not show any considerable differences. Also, as was shown in the previous sub-section the SIHC exhibits higher average annual benefit incomes than the LDS possibly reflecting an under-representation of recipients with low levels of benefit income in the SIHC.

Consistent with our other findings Figure 1 shows a considerable under-representation of the total benefit population in the SIHC when annual income is the reference point. From the figure it is also apparent that the SIHC significantly under-estimates the number of people on low levels of benefit income. This is consistent with our claim above with regard to the underreporting of those on benefits for short durations. However, what is also apparent is that

the SIHC under-represents the frequency of incomes at the peaks, which represent maximum benefit levels available. This seems to be mainly due to more noise in the household survey data, with greater frequencies of income around these peaks. Thus, it appears that benefit incomes represented in the SIHC are more volatile and skewed to an extent, favouring the upper middle range of benefit incomes and under-representing the number of people receiving low levels of benefit income over the year. The result is that long-term/higher annual benefit income recipients are over-represented in the observed benefit population in the SIHC.

If we look in more detail and examine annual benefit incomes by individual payment types we can see that this under-representing of recipients at the lower end of the distribution is especially apparent in payments with shorter average durations such as unemployment related benefits, sickness allowance and partner allowance (see Appendix 5).

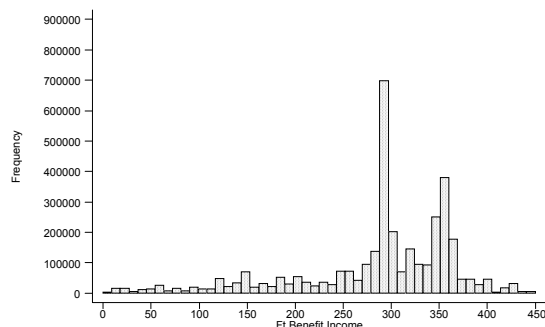
Figure 4.1: Distributions of annual benefit income, 1996/97



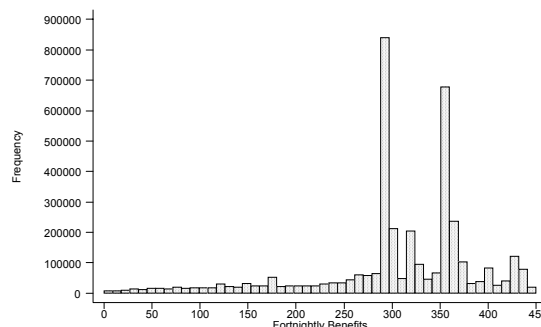
We now turn to comparison of the distribution of current income in the SIHC with fortnightly income in the LDS. As before we have doubled incomes reported in the SIHC to put them on the same basis as the LDS incomes. In Figure 4.2 we consider benefit income and in Figure 4.3 we consider total income. Appendix 6 shows distributions for particular payment types for benefit income alone while Appendix 7 reports distributions for payment types for total income.

Figure 4.2: Distributions of fortnightly benefit income, 1997/98

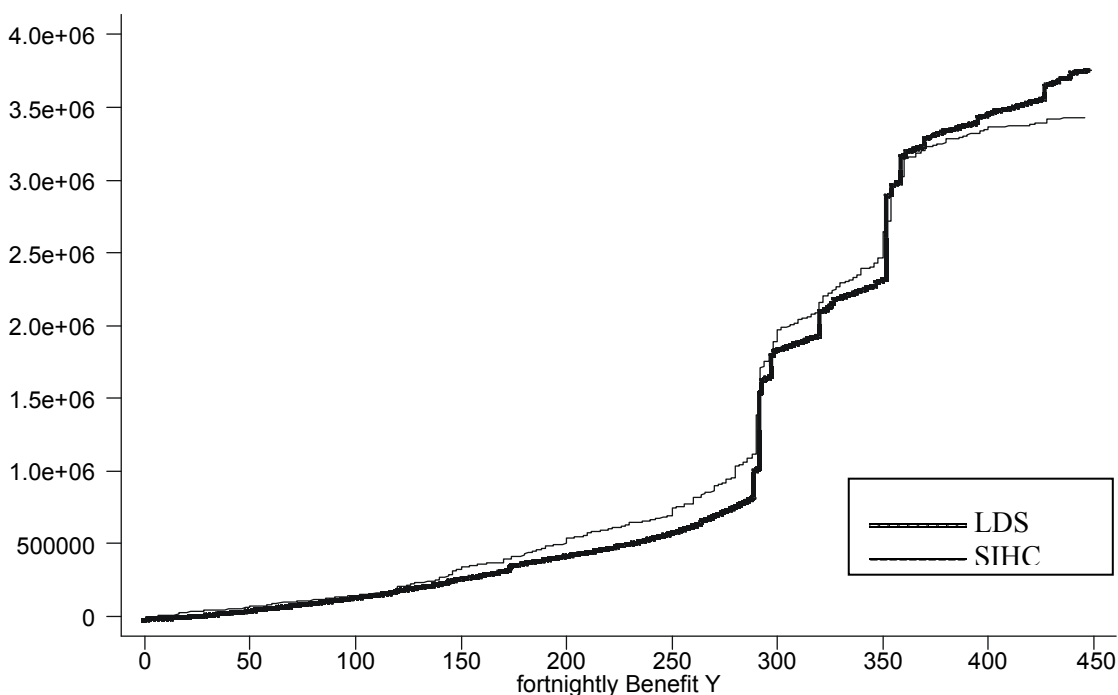
(a) Histogram of recipients in SIHC



(b) Histogram of recipients in LDS



(c) Cumulative distribution



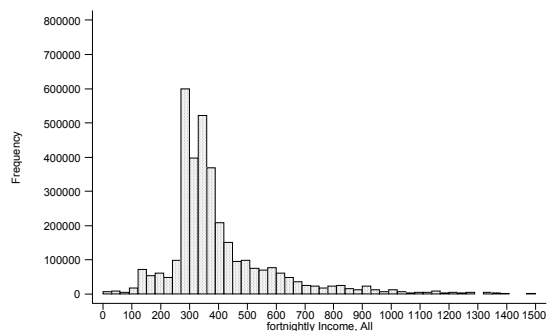
In contrast with the previous figure, Figure 4.2 shows no evidence of the under-representation of those with low benefit incomes. In fact the reverse is apparent. With regards to information on current income there does appear to be some evidence that a section of the observed benefit population are under-reporting their benefit incomes in the SIHC relative to the LDS at low to mid range fortnightly benefit levels (around \$150 a fortnight to just under \$300 a fortnight). There are also fewer people represented at the top of the distribution.

It is difficult (and probably not wise) to try and explain the differences in the two distributions. As the SIHC is a household survey standard errors are associated with the population estimates. The number of individuals in the tails of the income distribution is small thus the confidence interval is wide and differences across the data sources in the proportions of individuals at these tails is not likely to be statistically significant. Also, as we have noted, the robustness of these comparisons could be undermined if there is a consistent difference in the treatment of current weekly wage and salary income in SIHC and lump sum transfer payments. The problem of lump sums is not likely to emerge at the very low level of benefit payments, as lump sums would be both small and rare for recipients in this range of benefit.

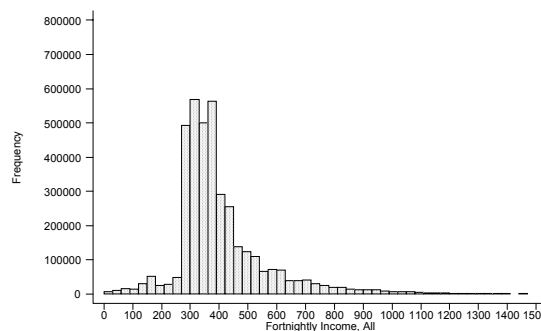
Similar patterns are observed among other payment types (if account is taken of the noise associated with some payments due to small sample sizes - see Appendix 6).

Figure 4.3: Distributions of total fortnightly income for benefit recipients, 1997/98

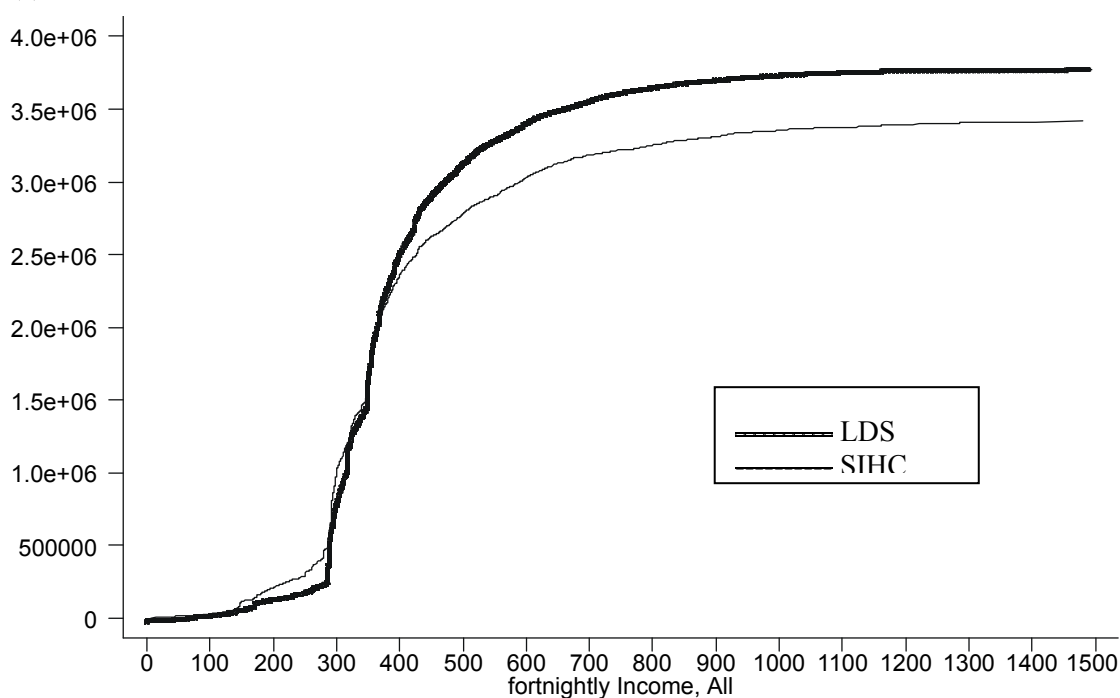
(a) Histogram of recipients in SIHC



(b) Histogram of recipients in LDS



(c) Cumulative distribution



The comparison in Figure 4.3 does not indicate much difference in total incomes for benefit recipients across the two data sets. The scale of the horizontal axis obscures some of the features but the comparison seems similar to that observed in Figure 4.2 with no under-reporting at very low levels of benefit (less than \$150 per fortnight), some evidence of under-reporting at \$150 to just under \$300 a fortnight with an increased proportion of recipients in this income range. Individuals with incomes in the top range of the income distribution are under-represented possibly reflecting the treatment of lump-sum advances. Due to means testing of benefits the low-income population is restricted to those with less than about \$500 of fortnightly income from earnings or other non-benefit income sources.

4.6 Key points

Key points which emerge from the comparison are:

- The primary problem with measuring incomes of the benefit recipient population is with non-reporting or under-sampling that occurs with certain particularly mobile sub-groups within the population – i.e. young and single people.

- Annual income measures understate those on benefit income for short periods and those who move between different categories of benefit payment throughout the year.
- There is strong evidence that a proportion of the benefit recipient population under-reports their annual benefit income when interviewed for the SIHC.
- It is difficult to determine whether under-reporting of current total or benefit income is apparent for self disclosed benefit recipients in the ABS survey. However total transfer incomes are under-stated in the SIHC, to an even greater extent than benefit population estimates, suggesting there is some evidence of under-reporting of incomes for the observed benefit population.

5. Reported vs imputed eligibility levels for income support recipients by payment type

In Section 4 we compared the characteristics of the benefit population from the SIHC with those of the benefit population derived from administrative data records. In this section we use another method to evaluate reported incomes of the benefit population. We ask the question: are those individuals in receipt of payments receiving what we expect them to receive given their reported characteristics?

Below we have taken Centrelink's requirements governing eligibility and income test arrangements for benefit payments and matched them with the reported demographic characteristics and private income details provided by the survey respondents in the SIHC. We then impute (a) what particular benefit (if any) each individual would likely be entitled to; and (b) the amount of entitlement after taking account of income test arrangements. We impute eligibility using the 1997/98 SIHC as the base population in the Melbourne Institute Tax and Transfer Simulator (MITTS) and use the system rules as of March 1998¹². Imputed entitlement levels are based on current private income levels observed for the base population. Although, entitlement to certain benefits may only be reviewed intermittently throughout any given year claimants are required to contact Centrelink if their circumstances change (financial and other) thus to capture a persons current circumstances the use of current income is appropriate.

5.1 Recipient numbers

Table 5.1 reports the results of the first part of this imputation. It compares (using the weights provided in the SIHC) the numbers of individuals one would expect to be eligible for specific major payment types with the number of individuals who report receiving these major payment types. Note that we only look at employees and those out of work. The complexity of the comparison is greatly expanded when one seeks to determine the payment status of the self-employed. The rows in the table represent the population estimate that MITTS imputes to be eligible for each major category of income support. The columns represent the population estimates reporting to be in receipt of the major categories of income support.

A majority of the observations lie on the diagonal, which shows that the imputations associated with the given structure of payments *align closely* with payment information reported by the respondents of the survey. However the table also shows that there are 1,318,958 individuals who appear to be eligible for payments who are not reporting receipt of any benefit at all. There are many reasons why this may be the case other than non-reporting. The observed characteristics of individuals in the data do not include all of the information required to determine eligibility for particular benefits. For instance information on assets is not available and thus some of this group may be ineligible for payment due to the assets test. Fortunately, the group of households that would not be eligible based on their level of assets (which excludes the home), but would be deemed eligible based on their level of income is relatively small. Particularly, because the SIHC records income from investments (like

¹² For details on MITTS and the imputation see Creedy et al (2002).

dividends or interest) and superannuation income, which are incorporated in the calculations, we do not feel that the lack of information on assets is likely to be a major problem.

Incomplete program participation (take-up) for certain individuals is also likely to be an important explanation for the discrepancy. Other requirements for eligibility, which we cannot observe, are whether someone has been a resident for at least two years and is actively seeking and willing to take on full-time employment. The latter may be particularly problematic.

To explore this issue further we now turn to examining the characteristics of those individuals who appear to be eligible for income support but claim not to be receiving any type of payment. In Table 5.2 we present characteristics such as age, gender, presence of investment income, country of birth, year of arrival, labour force status, marital status, educational qualification, area of residence, and number of dependent children of these individuals. We also separate out those with investment income in order to determine whether the assets tests may be a significant factor in explaining the discrepancy between imputed and reported eligibilities. As a comparison we also present the characteristics of the benefit population as observed in the survey, not including individuals only in receipt of family payment, and the general population. For convenience we will refer to this population as ‘non-reporters’, keeping in mind the range of explanations advocated for the divergence between imputed and reported outcomes.

Table 5.1: Reported vs. imputed eligibility, income support recipients by payment type^a

<i>Imputed benefit type</i>	<i>Reported benefit type</i>								<i>All^c</i>
	<i>Unemployed^b</i>	<i>Age/DVA Pension</i>	<i>Disability^c</i>	<i>Education^d</i>	<i>Parenting</i>	<i>Sole Parent Pension</i>	<i>Other benefit</i>	<i>Not on income support</i>	
Unemployed ^b	407,736	0	0	1,623	3,085	0	10,700	293,947	716,936
Age/DVA pension	0	1,851,353	47,899	373	0	0	11,872	203,957	2,063,378
Disability ^c	0	19,076	609,235	6,277	3,630	0	1,031	0	609,235
Education ^d	18,347	0	0	230,026	7,930	23,710	0	277,503	539,698
Parenting	17,247	0	0	4,359	412,466	1,929	1,414	296,944	733,294
Sole Parent Pension	2,192	0	0	9,609	3,504	283,482	472	54,179	344,761
Other benefit	116,515	24,181	4,253	7,311	4,711	1,244	210,609	192,428	544,808
Not on income support payment	55,557	39,527	6,121	47,607	45,119	2,957	12,650	-	208,696
Total	617,594	1,934,137	667,508	307,185	480,445	313,322	248,748	1,318,958	5,760,806

Notes: (a) Sample consists of the unemployed, those not in the labour force, and employees. (b) Unemployed includes NSA, JSA, MAA and Youth training allowance. (c) Disability includes DSP and sickness allowance. (d) Education includes Austudy and Abstudy. (e) Certain individuals in the SIHC report to be in receipt of a number of payment types thus the total in this column does not reflect the sum of individuals across payment types. The MITTS totals are not however subject to this as benefits are imputed on a mutually exclusive basis.

Table 5.2 shows that the young seem to be over-represented among the ranks of ‘non-reporters’. This is consistent with the result of our earlier comparison of benefit recipients with administrative data records: young recipients are under-represented in the ABS survey. We also observe that it is primarily those singles that have never married who are the non-reporters rather than those separated, widowed or divorced.

Table 5.2: Characteristics of individuals imputed to be eligible for payment but not reporting receipt of payment

	Imputed (per cent)	Benefit population (per cent)	Entire population (per cent)
15-24 yrs	34.0	11.8	18.0
25-44 yrs	32.8	26.0	39.5
45-64 yrs	20.3	24.0	27.9
65+ yrs	12.9	38.4	14.8
<i>Total</i>	<i>100</i>	<i>100</i>	<i>100</i>
Male	38.2	40.0	49.4
Female	61.8	60.0	50.6
<i>Total</i>	<i>100</i>	<i>100</i>	<i>100</i>
Income from investments	37.4	38.5	42.4
No income from investments	62.6	61.5	57.6
<i>Total</i>	<i>100</i>	<i>100</i>	<i>100</i>
Married/Defacto	48.3	54.1	59.9
Separated/Widowed/Divorced	11.1	25.6	12.8
Never married	40.7	20.4	27.3
<i>Total</i>	<i>100</i>	<i>100</i>	<i>100</i>
Oceania and Antarctica	74.3	72.2	75.9
Europe and former USSR	11.7	19.5	15.4
Asia	9.4	4.5	5.3
Other	4.6	3.9	3.3
<i>Total</i>	<i>100</i>	<i>100</i>	<i>100</i>
NA/Born in Australia	72.0	70.0	73.2
Arrived before 1976	8.9	19.4	14.1
Arrived 1976-1990	10.3	7.2	8.8
Arrived after 1990	8.9	3.5	4.0
<i>Total</i>	<i>100</i>	<i>100</i>	<i>100</i>
Employed full time	10.3	3.2	44.5
Employed Part time	18.7	8.9	14.5
Unemployed	9.8	11.6	5.3
Not in the labour force	61.1	76.3	35.6
<i>Total</i>	<i>100</i>	<i>100</i>	<i>100</i>
University qualification	10.3	3.7	12.1
Other qualification	20.9	24.0	28.9
No Qualifications	68.6	72.3	59.0
<i>Total</i>	<i>100</i>	<i>100</i>	<i>100</i>
No dependent children	54.7	70.2	60.7
Dependent children	45.3	29.8	39.4
<i>Total</i>	<i>100</i>	<i>100</i>	<i>100</i>
NA	2.6	1.4	2.4
Capital city	61.7	56.0	62.4
Rest of State	35.7	42.6	35.3
<i>Total</i>	<i>100</i>	<i>100</i>	<i>100</i>

Another general comment about the findings is that there appears to be no indication that the assets test-based exclusions are a major factor in explaining the difference between the observed and imputed benefit populations. Those with investment income are slightly under-represented in the non-reporting group relative to the general benefit population. While we

understand that this is a fairly crude measure, if assets testing were a major problem in the imputation you would expect to find a larger share of individuals with assets in the group not observed to be in receipt of benefits.

The employed, both part-time and full-time, are also strongly represented in the non-reporters group. This suggests there may be a 'take-up' issue: employed people are in receipt of other forms of income and thus may not feel the need to take-up their benefit, particularly if they are only entitled to a small amount. It is also possible that those working do not wish to disclose that they are in receipt of any benefit income.

University-educated people are also more likely to be a 'non-reporter' than other cohorts in the population. As the university educated population are on the face of it more employable, some attachment to the labour market is expected and thus this may also be a take-up issue if working part-time or casually. Stigma associated with receiving benefits may also be higher among the ranks of the highly educated. The stigma effect could also work in another way. The highly educated person may actually take-up the benefit that is available to them, but be reluctant to disclose this income. Note that these assertions are purely speculative

5.2 Entitlement levels

Now we turn to the difference between imputed and reported entitlement levels. If the difference is small for most individuals, we may simply be encountering a small measurement error and thus this issue would not be significant. However it also may highlight a take-up issue, as those eligible for small amounts of benefit perceive that the cost of taking up the benefit outweighs the benefit. Figure 5.1 shows the distribution of the absolute difference between reported benefits received and imputed benefits received for the estimated population either reported to be in receipt of a benefit or imputed to be in receipt of a benefit. Note that the sample does not include the self-employed.

The data presented below reveals that a large majority of the population have either no difference in entitlement or a very small difference. There is certainly a weighting to the left side of the graph indicated that the imputed number of recipients is greater than the number of reported recipients in more cases than vice-versa.

We now disaggregate the data set by payment type to see if particular payments are more at risk of apparent under-reporting of benefit income than others. As Figure 5.2 below shows, unemployment benefits and sole-parent pensions are the two categories of payment that are under-reported most.

Another significant pattern in Figure 5.2 is the spike in the level of parenting payment between \$30 and \$40 a week. This seems to highlight a take-up issue, with a lower percentage of those eligible for the basic payment only choosing to take up the benefit. There is no clear pattern evident in the data related to education payments. More detailed work in this area is required to shed more light on trends among these recipients.

Figure 5.1: Difference in benefit incomes (reported-imputed)

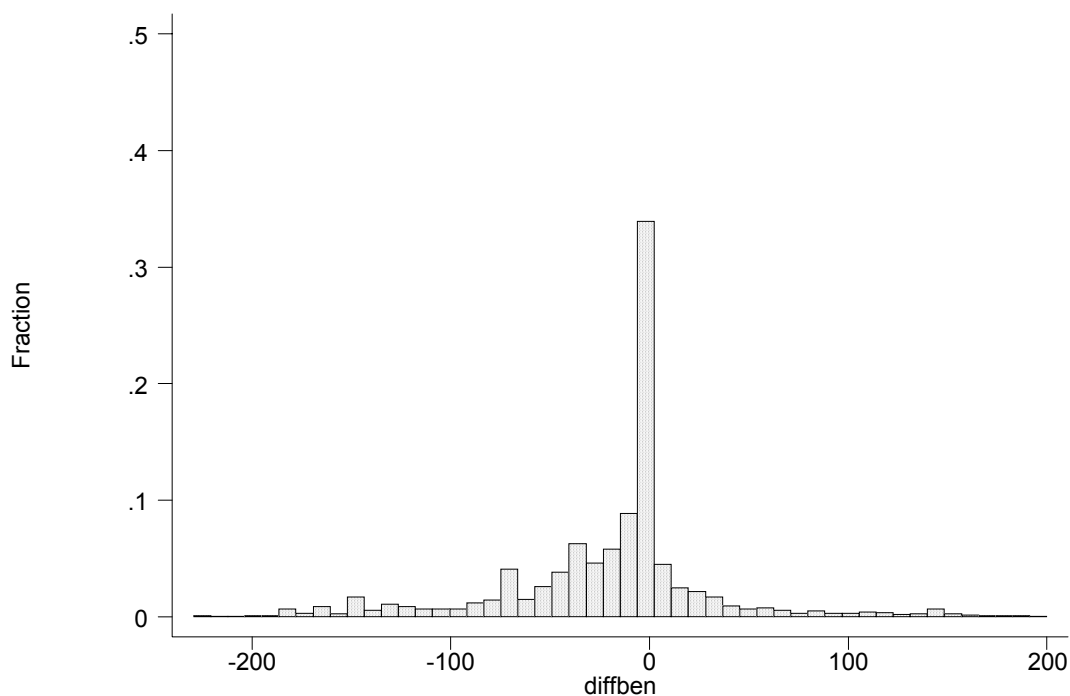
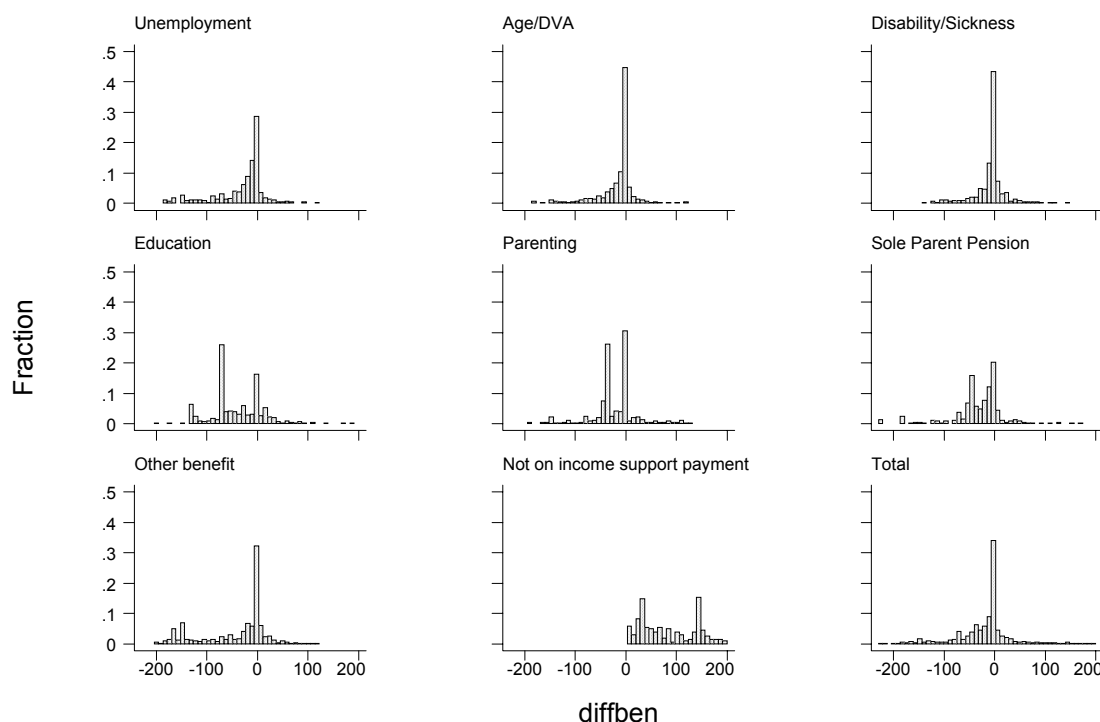


Figure 5.2: Difference in benefit income by imputed payment type (reported-imputed)



5.2 Income distribution

How does the distribution of income compare between reported net income levels and imputed net income levels? We know that MITTS over predicts the number of benefit recipients due to unobserved information relating to eligibility requirements. From the previous section we also know that the SIHC under-represents the benefit population. Thus,

the *true* income distribution will lie somewhere in between the two sets of results presented here.

So what do average net income levels look like if we believed the imputation procedure was correct, and how do these levels compare to the averages obtained from reported information on incomes in the base data set? A comparison of various summary measures of the income distribution across income units is presented in Table 5.3. Incomes are not equivalised and again we focus purely on employees and those not working. Average net incomes are presented along with a measure of inequality, the Gini coefficient, and a relative poverty measure, the headcount ratio. Other inequality measures and poverty measures were examined with the same emerging patterns.

Table 5.3: Comparison of reported and imputed net income unit income

	Reported income	Imputed income
Average incomes	525.04	533.80
Standard deviation	416.23	392.37
Gini coefficient	0.3857	0.3569
Relative poverty line ^a	213.50	218.54
Headcount poverty ratio	21.17	19.21

a) The relative poverty line was estimated using half of median income-unit income.

If the imputed data are correct, average net-incomes are significantly higher than are observed in the data. However, we know that the imputation process is likely to lead to an over-representation of the benefit population. Thus, if we believe that the survey respondents report their true levels of non-benefit income, average income unit incomes across the non-working and employees after taxes and transfers lie somewhere in between \$525 and \$534 a week. This increase in average incomes has an affect on measures of inequality and poverty as can be seen in the table. Using the MITTS imputations of net income, the Gini coefficient is around 7.5 per cent lower than that based on the net incomes observed in the data. The under-representation of the benefit population and the under-stating of benefit incomes in the SIHC also affect poverty measures. Although the relative nature of the measure leads to an increase in the poverty line, estimated using half median income-unit income, under-stating the incomes of the benefit population leads to as large as a two percentage point increase in the proportion of income units in relative poverty.

More detail on the relative positions of income units in the income distributions generated is provided in Table 5.4. The rows represent the income deciles using the reported net income distribution and the columns represent income deciles using the imputed net income distribution. MITTS does not model tax deductions thus we expect some differences at the middle to top end of the income distribution however we do not wish to focus on these. What we do wish to focus on is in what is happening to the relative positions of income units in the lower end of the income distribution. The table shows that 37.1 per cent of the income units in the bottom income decile move to higher income deciles when incomes are imputed based on observed characteristics. This has an effect on the second decile also, with 40.8 per cent moving to the first income decile when using the imputed income distribution. A similar pattern occurs higher up the distribution with most of the movement coming from income units moving up or down one decile in the distribution, however the tendency for this to occur decreases as the income decile increases. This highlights how important it is to correctly estimate incomes of the benefits population as relative positions in the income distribution can vary quite considerably.

Table 5.4: Position in income distribution, reported incomes vs. imputed incomes

Deciles of reported net incomes	Deciles of imputed net incomes									
	1	2	3	4	5	6	7	8	9	10
1	62.9	18.2	9.2	3.6	4.2	1.0	0.4	0.3	0.3	0.0
2	40.8	50.9	5.6	1.5	0.7	0.5	0.1	0.0	0.1	0.0
3	3.7	19.7	58.1	11.2	4.8	2.1	0.4	0.0	0.0	0.0
4	1.1	1.2	21.7	62.8	8.7	3.0	1.1	0.2	0.2	0.0
5	0.1	0.6	3.3	19.3	65.6	8.6	1.8	0.3	0.1	0.3
6	0.0	0.0	1.1	2.1	14.8	69.5	11.1	1.2	0.2	0.0
7	0.0	0.2	0.3	0.3	1.2	13.9	73.8	9.9	0.5	0.1
8	0.0	0.0	0.1	0.2	0.5	1.1	10.2	82.5	5.2	0.2
9	0.0	0.0	0.0	0.0	0.0	0.1	0.9	5.6	89.5	3.9
10	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	4.1	95.7

Note: the rows above sum to one hundred.

This section shows that under-representing the benefit population has an affect on the distribution of net income and can have quite a significant effect on measures of inequality and poverty. It highlights the importance of obtaining the correct population estimates of benefit income and getting the benefit population right.

6. Concluding comments and implications

The development of effective policy in regards to poverty demands accurate and timely information about economic and social conditions. However the process of measuring changes in conditions is subject to several sources of confusion. These include problems of interpretation, where the individual researchers own experience and worldview colours their perspective on data. Researchers thus draw different conclusions from the same data set on the basis of different personal backgrounds and biases. Another issue that arises concerns problems of definition, where different researchers use the same terminology to refer to different concepts. There are often significant variations in researchers understanding of core concepts such as income and wellbeing. Researchers also differ in their preference for using relative and absolute conceptions of poverty. A third issue and the main focus of this report concerns problems with the data used in analysis. Data difficulties can be divided into non-sampling errors, where the problem derives from the limited size of the sample chosen relative to the size of the total population being surveyed, and sampling errors, which stem from the character of the observations that make up the data set. Sampling errors grow as the complexity of measures derived from the sample data grows. Standard errors are used to quantify the extent of sampling error.

The main object of this report is to evaluate the data sources, particularly data from sample surveys, available for measuring the situation particularly of those with low income, and develop strategies to overcome any deficiencies.

6.1 Are the data relating to low incomes plausible?

Standard errors

In Section 2 we considered the statistical properties of the sample data and in particular errors associated with the performance of income and expenditure surveys. We found there are likely to be large standard errors attached to estimates using some data from the surveys. The problems become particularly acute for groups that make up very small proportions of the sample population.

Income vs expenditure

In Section 3 we undertook an analysis comparing the income and expenditure data from the Household Expenditure Survey (HES). We sought to understand the differences between those households in the bottom quintile of the income distribution and bottom quintile of the expenditure distribution. Those households that were represented in both the lowest income and expenditure quintiles we took to be the most impoverished households. We sought to explain the presence of households who appeared in the lowest income quintile but did not appear in the lowest expenditure quintile and posed the question of how many of these households conformed to expectations about the characteristics of conventionally defined disadvantaged households.

The compelling observation about these two quintiles was the high incidence of apparent dissaving among the low-income quintile. A proportion of this group were not represented in the low-expenditure quintile, reporting expenditures in the middle and upper quintiles of the distribution. A plausible explanation for this finding is that a significant proportion of these households were households in transition, with a head of household moving between jobs, beginning retirement and experiencing other periods of change. The evidence for this conclusion was that those households with zero or negative income did not exhibit the characteristics of typical poor households. They did not report financial distress much – a lower proportion went without meals and sought financial help from friends, and a negligible proportion sought help from welfare organizations.

Another explanation for the apparent high incidence of dissaving among the low-income quintile is that there is some underreporting of income occurring in the survey.

A third explanation is that the gap between income and expenditure occurs because HES does not record the running down of assets to finance expenditure or irregular receipts of income, such as a superannuation payout or the sale of a large durable item. A significant proportion of the bottom income quintile are older people who are quite understandably running down their assets over time to maintain their accustomed expenditure levels.

In sum, there are three key findings from this section. The first is that households with zero or negative income do not exhibit the characteristics of conventionally poor households. It seems likely that this group does indeed largely represent households in transitional situations and should be left out of conventional poverty measurement. Second, the category of households that were in lowest quintile on the basis of their income but not in the lowest quintile on the basis of their expenditure seems to contain a majority of older people who were legitimately running down their assets in order to maintain a higher standard of living. Some households in this category are undoubtedly experiencing poverty but this a small fraction of the total. Third, that category of households, which fall into the lowest quintile on the basis of expenditure but not on the basis of income, appears to be mainly a group with an above average propensity to save. Their tendency to spend at below average rates does not appear to be driven by financial distress in the manner expected of a low income, impoverished household. This may mainly be due to households with negative expenditure remaining in our population of low-expenditure households, thus further research needs to be undertaken to evaluate the effect of their removal from the sample.

Survey data vs administrative data

In Section 4 we undertook a comparison of the ABS's income and expenditure survey data and the administrative records of transfer payments made by the Department of Family and Community Services and its predecessors. We found evidence of discrepancies between the survey data and the administrative data and several underlying problems with surveying methods that may explain these discrepancies.

The SIHC under-estimates the benefit population, with certain benefit types more at risk of being under-estimated. Unmarried people and those in de facto relationships are quite significantly under-represented across all the years examined. The young seem under-represented as well, especially those under 25 years. Geographic mobility is one trait shared

by these two groups and a case may be made that the ABS's survey is not able to locate and represent these people in its sample. Shorter-term receipt of benefits is also common for these two groups and this may mean that people do not see the need, or forget, to report their brief experience with receiving benefits. Newstart payments are typically received for a short time and this group is also under-represented in the ABS data.

Annual income measures seem to understate those on benefit income for short periods and those who move between different categories of benefit throughout the year. Examination of the finer detail of individual payments for those receiving benefits for short periods (mainly Newstart, Youth Training Allowance, Partner Allowance and Sickness Allowance) reveals that this group appears to be most susceptible to having higher average incomes. As well, those individuals who move between Mature Age Allowance and Disability Support Pension appear over-represented in high average income cohorts. Confusion between these benefits may drive the average up making it difficult to draw conclusions about the actual trend being observed.

Key points which emerge from the comparison are:

- The primary problem with measuring incomes of the benefit recipient population is with non-reporting or under-sampling that occurs with certain particularly mobile sub-groups within the population – i.e. young and single people.
- Annual income measures understate those on benefit income for short periods and those who move between different categories of benefit payment throughout the year.
- There is strong evidence that a proportion of the benefit recipient population under-reports their annual benefit income when interviewed for the SIHC.
- It is difficult to determine whether under-reporting of current total or benefit income is apparent for benefit recipients. However total transfer incomes are under-stated in the SIHC, to an even greater extent than benefit population estimates suggesting there is some evidence of under-reporting of incomes for the observed benefit population.

Imputed benefit eligibility

In Section 5 we compared the income of the actual benefit population with the income profile we would expect given their reported characteristics and eligibility rules applied by Centrelink.

We found significant alignment between the ABS survey and our imputed income profile. However approximately 1.3 million individuals appear to be eligible for payments but do not report receiving them. Much of this is probably failure to report from the survey respondent but there are clearly other possible explanations including factors that may disqualify people from benefits, such as information on whether individuals are conforming or not to the unemployment benefits' jobseeking activity test. Another issue is take-up; some benefits are available to individuals who are employed, but there is probably a disinclination to take these benefits if they are only a small amount or if they have a burden of investigation and paperwork for the individual. The popular stigma against admitting to or even taking up a "welfare" payment may also add to the fluctuations and gaps in the data.

This section also showed that under-reporting of incomes of the benefit population can have significant effects on measures of inequality and poverty.

6.2 Overall interpretation of results

We considered the statistical properties of the survey data noting the limitations on inference as a result of the error bounds. We then attend to the plausibility of the data in three main ways. First, we consider the situation of those reporting low income and expenditure in household surveys. Second we compare the circumstance of those reliant on government pensions and benefits in the surveys with administrative data. Finally we compare the nature of pension and benefit recipients in the surveys with what might be expected from eligibility criteria.

We found that the low-income and expenditure populations were heterogeneous and only a part of them had the characteristics of the conventionally disadvantaged. Hence we should be very wary about basing welfare policy on aggregate information from these populations. However subsections of the populations may conform to conventional views of the needy. Significant under-reporting income is a plausible explanation for implausibly high gaps between income and expenditure among low-income people. However for other households in the low-income and expenditure groups there are other plausible explanations.

Many of the families and people in the lowest income quintile from the income surveys are people in transition, and other households are likely to be in situations where assets and savings are being depleted to maintain higher levels of consumption with no welfare implications. These situations are not necessarily of concern to policy makers worried about disadvantage and distress.

The comparison of the survey data with the administrative data also revealed important differences. These differences may be apportioned between (i) limitations in survey coverage and in under-reporting (ii) problems with the administrative database stemming from inaccurate recording, missing observations and so on. The comparison suggests there may be significant underreporting among recipients of transfer payments. However much of the underreporting may be from individuals who receive only small infrequent or temporary payments. These situations of underreporting are among groups who are less likely to be of concern to welfare policy analysis. The underreporting is not evenly spread across the welfare population. Current income is more accurately reported than annual income. There is wide variation in reporting between payment types with greatest discrepancies among unemployment benefit recipients and least among old age pensioners.

The comparison of survey data and estimates based on entitlement also noted large discrepancies suggesting either problem with surveys, limitations of the entitlement calculations or failure of eligible people to take-up pensions. Two problems arise from the survey sample; failure to cover the appropriate population and under-reporting of income. The calculations of entitlement are complex and it is likely that failure to account for some criteria such as asset and activity tests will lead to overestimation of the survey population. There is also the prospect that take-up will be incomplete particularly where only small amounts of payment are involved. Nevertheless the comparison suggests under reporting of benefit. These results confirm observations from overseas studies (see Behrendt, 2000).

Overall the analysis suggests that we ought to be wary of low-income policy based on survey data alone but that under-reporting is only one of a number of explanations for apparent contradictions. This suggests that it is neither sensible to place too much emphasis on survey conclusions nor to ignore the findings of the surveys. It is likely that how much emphasis can be placed on the data will depend on the particular circumstance of its use. In some situations the survey data will provide a useful basis for suggesting policy, in other circumstances it will not. A case in point is the calculation of the apparent level and changes in poverty. Problems may occur in the way in which survey data is used, the assumptions used in measuring poverty and the influences drawn from such data. It would be useful to undertake research specifically looking at the way limitations of data interact with other matters in relation to poverty analysis.

6.3 Strategies for dealing with data quality concerns

The discussion so far has outlined problems with data quality and likely causes. It remains to outline a strategy for dealing with them. It seems inevitable that there will always be issues of data quality but useful results and policy will also continue to be drawn from existing surveys and other sources of data. The following list outlines the strategy the researcher needs to adopt or be mindful of, in presenting results in writing reports:

- Triangulate results
- Undertake sensitivity analysis
- Decompose and disaggregate

- Clarify definitions
- Provide full information (eg standard errors)
- Note limitations in discussion
- Note alternative interpretations

The discussion that follows is related to research reports rather than academic papers. In many journals there is not sufficient space to qualify many of the findings and much of what is outlined below may be severely abbreviated or just left out (with the understanding that those reading the paper will know about them). However for reports written for policy makers where space is less of a constraint, the detail is required.

Triangulation

While results from one suspect source may be of doubtful value, when several independent suspect sources produce similar results then confidence in the overall result rises. Consider for instance the issue of inequality. In recent times measures of inequality of household income, measured by disposable income utilising income and expenditure surveys have suggested a widening gap during the eighties, continuing during the early nineties but at a lower rate and virtually constant in the late nineties (see for instance Johnson and Wilkins, 2002). Both income and expenditure surveys have suffered from changes in definitions and concerns about high standard errors. There are implausibly high marginal propensities to consume at low-income levels. Aggregation of household results also identifies in the order of ninety percent of income and expenditure reported in national accounts. The conclusion in regard to rising inequality (since the early eighties) might be regarded with some suspicion considered on its own. However there is other supporting evidence. Wage data also reveal a widening trend over this period. Other information suggests that wage rates for higher paid jobs and more skilled jobs have also grown at a faster rate than lower paid jobs. Restrictions on wage setting in the labour market have been removed leading to the possibility of increased dispersion in wage rates. Measures of executive salaries suggest much faster increases than average and below average salaries. Rates of payment for social security transfers have risen at the same rate as average weekly earnings (the old age pension is tied to average weekly earnings). Returns to factors of production have moved in favour of capital and against labour suggesting faster growth of unearned income that would more typically go to high-income earners. Most of this ancillary evidence supports the proposition that household incomes are likely to have become more unequal over the period since the early eighties. While resort to the income and expenditure surveys alone might cause the researcher to be tentative the supporting evidence suggest a more robust conclusion.

Sensitivity analysis

A time-honoured tactic when in doubt about a particular assumption is to consider a range of assumptions covering the spectrum of feasible options. Thus instead of presenting a single result in relation to the example of inequality mentioned above, the researcher would consider a range of results utilising a range of assumptions. In poverty research, it would be appropriate to report results using a range of equivalence scales, a range of income definitions, methods which generated both absolute and relative measures of poverty and a range of indexes of poverty corresponding to different levels of poverty aversion.

Producing a range of results may be a mixed blessing from a policy point of view. Often policy makers want a single best-bet result. However knowledge of the effect on assumptions on the results is invaluable to the researcher. It is often possible to attach probability bounds to particular assumptions and generate a probability weighted result.

Decomposition and disaggregation

Decomposition and disaggregation is often a required component of research since central interest may be focussed on the circumstance of particular groups. However even when this is not the case decomposition can be useful as a plausibility check. There is frequently other evidence (perhaps anecdotal) for particular groups and the disaggregated results may be

checked for plausibility against this other evidence. Alternatively decomposition can highlight the plausibility of the anecdotal based position put by lobby groups for particular positions.

Definitions

Whatever results are presented it is always incumbent on the researcher to fully define and articulate concepts, measures and indexes used in analysis. As has been discussed, many supposed differences in results between studies arise from the use of different definitions. However if definitions are not explicitly made, tracking them down to clarify the source of alleged differences can be tedious.

Full information

It is not always feasible to provide error bounds for results but some appreciation of the errors implicit in the methods is highly desirable. In general, it has not always been accepted practice to report standard errors in studies of poverty, inequality and social wellbeing in Australia. Admittedly there is a danger of unnecessarily clogging up results and, as the discussion has mentioned, statistical standard errors are themselves, subjective. The convention of describing results with a less than one in twenty chance of being random as robust is just that – a convention. In many policy circumstances a less stringent test (one in ten) may be appropriate. Explaining all of this in a report may lead to confusion. However such information can be put in footnotes or appendices and not necessarily interrupt the flow of the prose.

Noting limitations

The general rule should be to note all limitations of which the researcher is aware. These may be limitations of concept, of definition and of data. The limitations should not weaken the flow of argument and some craft may be necessary to include them in the structure of a report.

Note alternative approaches

Finally a well-rounded report ought to refer to alternative approaches. Generally reports will contain a literature review that describes other studies. Included in these will be studies that have taken an alternative, perhaps competing, perhaps complimentary approach. The relationship between the present study and these alternatives should be noted and explored.

6.4 Further work

The report has detailed inconsistencies in the data on low-income families and individuals from a variety of sources. The sources are the income and expenditure surveys, the administrative database, and the calculation of benefits according to entitlement within the MITT's model.

The report has identified a number of plausible explanations that reconcile the results from the surveys. However policy analysis requires more specific knowledge about the way in which data quality impinges on research outcomes.

It is not possible to make general deductions about this. The crucial matter is the way in which data are used. For some policy questions, the limitations will not be important, for others they will be crucial. Consider the specific example of the effect of the data limitation on measurement of poverty.

Estimations of poverty have been strongly criticised on the basis that the data on which they are based is unreliable. It is true that for some purposes the data are unreliable but for others the limitations are not important. How important they are depends crucially on the research questions being addressed. It would be desirable to undertake a further research project where the matter of the limitation is related to a specific research question and could be explicitly addressed. Given the importance attached to poverty measurement and the debate that occurred early this year (Smith Family, 2001, 2002; CIS 2001, 2002) a particularly useful research project would be to investigate the affect of data limitations for the measurement of

poverty. An outline of a suggested project and some thoughts about how it might be undertaken is provided in Appendix 8.

References

ABS 1994, *National Income, Expenditure and Product*, Catalogue No. 5206.0, June Canberra, AGPS

ABS 2001, *Income distribution, Australia 1999-2000*, Catalogue No 6523.0, February, Canberra.

ABS 2002, "Upgrading household income distribution statistics", *Australian Economic Indicators*, Catalogue No. 1350.0, April, Canberra.

Atkinson, A.B, Brandolini, A and Smeeding, T.M 2001, "Producing time series data for income distributionL sources, methods and techniques", *LIS Working paper No 295*, January, 37pp

Barrett, G, Crossley, T and Worswick, C 2000, 'Consumption and Income Inequality in Australia', *Economic Record*, Vol 76, No 233, June, 116-138

Behrendt, C (2000) 'Is there income poverty in Western Europe? Methodological Pitfalls in the Measurement of Poverty in a Comparative Perspective', Luxembourg Income Study Working Paper No. 258

Blacklow, P and Ray, R 2000, 'A Comparison of Income and Expenditure Inequality Estimates: the Australian Evidence, 1975-6 to 1993-4', *The Australian Economic Review*, Volume 33, No 4, December, 317-329

Bradbury, B 2002a Australia 1985-6 Documentation, obtained from the www.lisproject.org go to LIS and click on Australia, then 1985-6

Bradbury, B 2002b Australia 1981-2 Documentation, obtained from the www.lisproject.org go to LIS and click on Australia, then 1981-2

Burkhauser, R.V and Smeeding, T.M 2001, 'The Role of Micro-level Panel Data in Policy Research', paper presented to a seminar at the Melbourne Institute of Applied Economic and Social Research, Melbourne, 27 March

Carnaghan, M 1998, Does demand create poor quality supply: a critique of alternative distributional analyses, mimeo, Department of the Treasury, Canberra.

Creedy, J, Duncan, A., Harris, M., and Scutella, R. 2002, *Microsimulation Modelling of Taxation and the Labour Market: the Melbourne Institute Tax and Transfer Simulator*, Edward Elgar Publishing Ltd, United Kingdom.

Frosztega, M and the Households Below Average Income team 2000, Comparisons of income data between the Family Expenditure Survey and the Family Resources Survey, GSS methodology Series No 18, Government Statistical Service, London

Harding, A 1997, The Suffering Middle: Trends in Income Inequality in Australia 1982 to 1993-4, *Australian Economic Review*, 30, 341-358

Harding, A and Greenwell, H 2002, Trends in Income and Consumption Inequality in Australia, Paper prepared for the 27th General Conference of The International Association for Research in Income and Wealth, Stockholm, Sweden, August 18-24.

Johnson, D., Manning, I and Hellwig, O 1995, *Trends in the Distribution of Cash Income and Non-cash Benefits*, Canberra: AGPS

Johnson, D., Manning, I and Hellwig, O 1998, Trends in the Distribution of Income in Australia, *Australian Journal of Labour Economics*, Vol 2 No 1, March, 1-28

Johnson, D. T and Wilkins, R.W 2002, 'The effects of changes in family composition and employment patterns on the distribution of income in Australia', Paper presented to the Towards Opportunity and Prosperity conference, The University of Melbourne, April

Kroon, H and McDonald, M 1998 Robustness Assessment Report on Survey of Income and Housing Costs (SIHC) for 1995-6, downloaded from: www.lisproject.org and click on LIS, Australia; 1994 in that order

Luxembourg Income Study (2002) see website at www.lisproject.org

Moser, C.A and Kalton, G (1971) *Survey Methods in Social Investigation*, London: Heinemann

Social Policy Research Centre 2001a *Validation of IDS/SIHC data*, preliminary unpublished report BP/ 01-3, obtained from Bruce Bradbury

Social Policy Research Centre 2001b *Comparisons of Current and Annual Income Distributions*, preliminary unpublished report BP/ 01-4, obtained from Bruce Bradbury

The Canberra Group 2001, *Expert group on household income statistics, Final report and recommendations*, Ottawa available from www.lisprojects.org/links.html and click on Canberra group button

Travers, P and Richardson, S. (1993) *Living Decently*, Melbourne: Oxford University. Press.

Tseng, Y. and Wilkins, R. (2002), *Reliance on income support in Australia*, Melbourne Institute Working Paper No. 6/02, The University of Melbourne.

The Poverty Group 2001 "Looking Beyond Averages: A Research program on Poverty and Inequality", The World Bank, November 6.

Appendices

Appendix 1 Comparison of low income and expenditure households (raw data)

Table A1 Comparison of households in the bottom income quintile, the bottom total expenditure quintile and in the HES sample as a whole, 1998/99¹

Characteristic	Zero or Neg. income	Bottom quintile ² Income	Bottom quintile ² Expenditure	Whole sample ²
Household family composition, %				
Lone person	42.0	66.1	60.3	24.0
Couple only	26.3	18.8	18.7	24.6
Couple with dependent children	18.7	4.8	11.9	33.5
One parent with dependent children	3.3	8.7	5.6	8.5
Other	9.7	1.7	3.5	9.5
Gender of household reference head, %				
Male	46.6	39.7	56.2	61.0
Female	53.4	60.3	43.8	39.0
Age of household reference head, %				
Less than 25 years	4.2	4.1	3.6	5.4
25 to 64 years	79.6	48.6	62.4	75.5
65 years or older	16.3	47.4	34.1	19.2
Presence of dependent children (%)				
Dependents present	20.1	12.3	13.3	35.9
Labour force status (%) (Household reference head)				
Wage and salary – full-time	11.8	1.9	38.2	49.1
Wage and salary – part-time	13.0	6.2	5.1	9.4
Self-employed	34.9	6.4	5.8	7.3
Unemployed	2.7	8.7	4.5	2.8
Not in the labour force	37.7	76.8	46.4	31.5
Principal source of income (%)				
Government benefits		78.9	45.3	28.6
Wage and salary		6.3	40.8	57.5
Self-employed		3.3	6.0	6.3
Superannuation, investment, other		11.2	7.9	7.5
Income and expenditure				
Total household expenditure (\$pw)	602.8	329.0	111.3	519.8
Household disposable income (\$ pw)	-468.3	214.8	808.9	896.8
Total household income (\$pw)	-480.1	210.5	609.0	716.8
Cash flow problems (%)				
Could not pay electricity/gas/telephone bill	13.7	19.0	11.9	16.1
Could not pay registration/insurance	9.5	6.3	3.4	6.5
Went without meals	1.7	6.8	4.1	2.7
Unable to heat home	0.0	4.9	3.3	2.3
Sought assistance from welfare organisation	0.1	6.4	4.1	3.5
Sought financial help from friends	5.5	11.7	6.2	10.0
Had cash flow problems in past year	19.5	26.0	16.6	22.1

Note: Income quintiles are based on total household disposable incomes. Expenditure quintiles are based on total household expenditure on goods and services minus expenditure on income tax.

Table A2 Characteristics of lowest income but not expenditure quintiles and vice versa^{1 2}

Characteristic	Lowest income but not expenditure quintile	Lowest expenditure but not income quintile	Low income and low expenditure quintile
Expenditure/Income quintile			
Second	49.4	24.3	-
Third	24.9	24.3	-
Fourth and fifth	25.8	51.3	-
Household type			
Person living alone	49.8	39.3	86.7
Couple only	27.0	26.9	8.4
Couple with children	8.5	21.2	0.1
Lone parent	12.6	7.1	3.7
Other	2.1	5.4	1.1
Other demographic			
Female	57.9	28.3	63.2
Dependents under 25	19.4	21.2	3.2
Retired persons in hh	42.3	15.7	62.3
Age of head			
Under 25 years	4.8	3.9	3.1
25 to 65 years	57.3	82.0	37.6
65 years plus	37.9	14.0	59.3
Marital status of head			
Never married	17.4	26.1	21.3
Widowed/divorced/separated	43.8	23.8	68.8
Married/defacto	38.8	50.1	9.9
Principal source of income			
Wages and Salaries	9.6	71.5	2.2
Self employed	4.8	9.6	1.5
Other private income	14.9	9.0	6.5
Benefits	70.3	9.9	89.9
Labour force status of head			
Full-time employee	2.8	68.0	0.7
Part-time employee	8.8	6.8	3.0
Self employed	9.2	8.1	2.9
Unemployed	8.1	0.7	9.3
Not in labour force	71.1	16.5	84.1
Average weekly income/expenditure			
Expenditure	462.9	72.8	159.7
Disposable Income	215.8	931.0	203.7
Financial distress			
Could not pay gas/electricity/phone bill	20.8	8.1	16.7
Could not pay registration/insurance	8.8	3.6	3.1
Went without meals	6.6	1.8	7.1
Unable to heat home	4.7	1.7	5.3
Sought assistance from welfare organisation	5.0	0.8	8.2
Sought financial help from friends	14.0	4.2	8.6
Had cash flow problems in past year ³	28.2	11.3	23.3
Number of households in sample	724	748	538

Notes: (i) Income quintiles are based on total household disposable incomes. Expenditure quintiles are based on total household expenditure on goods and services minus expenditure on income tax. (ii) Households with zero and negative incomes have been omitted. (iii) "Had cash flow problems in past year" is a composite variable derived from the outlined financial distress variables

Appendix 2: Comparison of estimates of recipient numbers using survey data and administrative data**Table A3: Weekly/fortnightly recipient numbers by payment type**

	SIHC	LDS	Difference	Ratio
1996/97				
Total	3,405,728	3,637,812	-232,084	0.94
NSA/JSA	603,093	747,423	-144,330	0.81
Age Pension	1,571,442	1,639,427	-67,985	0.96
MAA	64,122	57,335	6,787	1.12
DSP	483,519	507,642	-24,123	0.95
Sole parent pension	302,545	341,712	-39,167	0.89
Wife/carer pension	172,215	160,285	11,930	1.07
Sickness allowance	59,560	18,788	40,772	3.17
Widow Allowance	46,809	50,158	-3,349	0.93
Special benefit	28,356	15,388	12,968	1.84
Partner allowance	60,096	68,888	-8,792	0.87
Youth training allowance	21,657	30,765	-9,108	0.70
1995/96				
Total	3,290,855	3,595,619	-304,764	0.92
NSA/JSA	668,709	751,481	-82,772	0.89
Age Pension	1,551,801	1,584,727	-32,926	0.98
MAA	44,606	55,454	-10,848	0.80
DSP	419,857	483,935	-64,078	0.87
Sole parent pension	252,148	326,473	-74,325	0.77
Wife/carer pension	113,675	177,265	-63,590	0.64
Sickness allowance	56,697	41,681	15,016	1.36
Widow Allowance	53,956	66,162	-12,206	0.82
Special benefit	23,109	19,135	3,974	1.21
Partner allowance	103,696	63,931	39,765	1.62
Youth training allowance	4,573	25,377	-20,804	0.18
1994/95				
Total	3,258,527	3,622,515	-363,988	0.90
NSA/JSA	658,499	731,838	-73,339	0.90
Age Pension	1,522,158	1,570,462	-48,304	0.97
MAA	52,825	49,192	3,633	1.07
DSP	399,148	447,031	-47,883	0.89
Sole parent pension	243,499	299,215	-55,716	0.81
Wife/carer pension	158,141	174,062	-15,921	0.91
Sickness allowance	51,337	40,215	11,122	1.28
Widow Allowance	58,660	65,538	-6,878	0.90
Special benefit	18,997	19,915	-918	0.95
Partner allowance	101,205	210,469	-109,264	0.48
Youth training allowance	2,218	14,577	-12,359	0.15

Table A4: Annual recipient numbers by payment type, 1995/96

	SIHC	LDS	Difference	Ratio
All benefits	3,525,337	4,549,300	-1,023,963	0.77
NSA/JSA	878,499	1,467,700	-589,201	0.60
Age Pension	1,436,251	1,688,900	-252,649	0.85
MAA	68,115	70,100	-1,985	0.97
DSP	461,556	541,200	-79,644	0.85
Sole parent pension	316,100	431,000	-114,900	0.73
Wife/carer pension	183,535	195,800	-12,265	0.94
Sickness allowance	85,453	122,100	-36,647	0.70
Widow Allowance	49,384	74,300	-24,916	0.66
Special benefit	31,824	51,300	-19,476	0.62
Partner allowance	65,564	96,600	-31,036	0.68
Youth training allowance	30,710	83,600	-52,890	0.37

Appendix 3: Comparison of estimates of expenditure on benefits using survey data and administrative data**Table A5: Total fortnightly benefit income by payment type, \$'000s**

	SIHC	LDS	Difference	Ratio
1996/97				
Total	973,538	1,156,137	-182,599	0.84
NSA/JSA	169,979	233,332	-63,353	0.73
Age Pension	435,465	507,207	-71,742	0.86
MAA	18,471	18,007	464	1.03
DSP	151,026	173,270	-22,244	0.87
Sole parent pension	94,468	124,286	-29,818	0.76
Wife/carer pension	45,303	46,444	-1,141	0.98
Sickness allowance	17,753	6,254	11,499	2.84
Widow Allowance	14,610	17,762	-3,152	0.82
Special benefit	7,407	4,959	2,448	1.49
Partner allowance	16,033	19,658	-3,625	0.82
Youth training allowance	3,018	4,957	-1,939	0.61
1995/96				
Total	913,713	1,114,271	-200,558	0.82
NSA/JSA	180,170	228,669	-48,499	0.79
Age Pension	424,632	479,028	-54,395	0.89
MAA	11,575	16,937	-5,362	0.68
DSP	126,455	159,841	-33,386	0.79
Sole parent pension	77,142	116,193	-39,050	0.66
Wife/carer pension	28,896	49,540	-20,644	0.58
Sickness allowance	15,944	13,772	2,172	1.16
Widow Allowance	15,778	23,083	-7,305	0.68
Special benefit	6,741	5,835	906	1.16
Partner allowance	25,568	17,671	7,897	1.45
Youth training allowance	810	3,702	-2,892	0.22
1994/95				
Total	880,055	1,078,090	-198,035	0.82
NSA/JSA	182,264	212,814	-30,550	0.86
Age Pension	397,751	457,798	-60,047	0.87
MAA	15,041	14,475	566	1.04
DSP	109,199	141,807	-32,608	0.77
Sole parent pension	74,058	103,620	-29,562	0.71
Wife/carer pension	38,558	46,695	-8,137	0.83
Sickness allowance	14,925	12,522	2,403	1.19
Widow Allowance	17,394	22,137	-4,742	0.79
Special benefit	4,884	5,859	-976	0.83
Partner allowance	25,708	58,455	-32,748	0.44
Youth training allowance	273	1,909	-1,636	0.14

Table A6: Total annual benefit income by payment type, 1995/96, \$'000s

	SIHC	LDS	Difference	Ratio
All benefits	22,337,699	27,856,780	5,519,081	0.80
NSA/JSA	3,858,717	5,716,728	1,858,011	0.67
Age Pension	9,924,669	11,975,693	2,051,024	0.83
MAA	453,558	423,432	-30,126	1.07
DSP	3,528,297	3,996,020	467,723	0.88
Sole parent pension	2,137,785	2,904,820	767,035	0.74
Wife/carer pension	1,120,018	1,238,498	118,480	0.90
Sickness allowance	393,738	344,310	-49,427	1.14
Widow Allowance	338,086	577,072	238,985	0.59
Special benefit	139,486	145,873	6,387	0.96
Partner allowance	372,906	441,782	68,877	0.84
Youth training allowance	70,441	92,552	22,111	0.76

*Appendix 4: Average incomes across individuals***Table A7: Average annual benefit incomes by payment type, 1995/96**

	SIHC		LDS	
	Mean	Standard deviation	Mean	Standard deviation
All benefits	6,336	2,724	6,123	3,014
NSA/JSA	4,392	2,706	3,895	2,708
Age Pension	6,910	2,329	7,090	2,496
MAA	6,659	2,079	6,040	2,589
DSP	7,644	2,084	7,384	2,596
Sole parent pension	6,763	2,738	6,740	3,282
Wife/carer pension	6,102	2,480	6,325	2,131
Sickness allowance	4,608	3,073	2,820	2,326
Widow Allowance	6,846	2,921	7,767	2,565
Special benefit	4,383	3,747	2,844	2,778
Partner allowance	5,688	2,179	4,573	2,429
Youth training allowance	2,294	1,509	1,107	1,156

Table A8: Average fortnightly benefit incomes by payment type

	SIHC		LDS	
	Mean	Standard deviation	Mean	Standard deviation
1996/97				
All benefits	285.85	81.15	305.59	81.96
NSA/JSA	281.85	76.76	300.17	77.64
Age Pension	277.11	85.29	297.48	82.69
MAA	288.06	56.38	301.99	49.32
DSP	312.35	66.27	328.20	70.31
Sole parent pension	312.24	79.98	349.73	86.97
Wife/carer pension	263.06	81.05	278.61	60.58
Sickness allowance	298.06	54.55	320.04	86.29
Widow Allowance	312.12	75.33	340.51	61.52
Special benefit	261.23	112.63	309.88	87.48
Partner allowance	266.80	52.06	274.39	65.45
Youth training allowance	139.36	70.19	154.93	70.40
1995/96				
All benefits	277.65	75.91	297.98	79.44
NSA/JSA	269.43	71.91	292.59	75.36
Age Pension	273.64	78.33	290.65	80.83
MAA	259.50	68.32	293.68	44.84
DSP	301.19	65.79	317.59	68.24
Sole parent pension	305.94	72.14	342.21	82.16
Wife/carer pension	254.20	87.27	268.72	61.35
Sickness allowance	281.22	64.86	317.72	71.88
Widow Allowance	292.42	82.58	335.47	58.89
Special benefit	291.72	72.09	293.21	84.66
Partner allowance	246.57	60.98	265.78	61.58
Youth training allowance	177.21	54.77	140.27	68.38
1994/95				
All benefits	270.08	79.83	286.16	77.24
NSA/JSA	276.79	92.31	279.61	76.54
Age Pension	261.31	81.82	280.29	79.00
MAA	284.73	44.04	282.93	41.78
DSP	273.58	68.99	305.02	65.90
Sole parent pension	304.14	59.10	332.99	76.31
Wife/carer pension	243.82	72.05	257.95	59.98
Sickness allowance	290.73	67.42	299.39	72.76
Widow Allowance	296.53	78.36	324.78	56.03
Special benefit	257.07	75.12	282.90	89.28
Partner allowance	254.01	50.73	267.06	65.12
Youth training allowance	123.24	49.58	125.91	64.10

Table A9: Average fortnightly total incomes by payment type

	SIHC		LDS	
	Mean	Standard deviation	Mean	Standard deviation
1996/97				
All benefits	400.03	194.28	403.64	157.01
NSA/JSA	387.94	218.49	342.47	110.89
Age Pension	371.73	139.89	400.50	130.85
MAA	348.29	272.13	351.33	91.65
DSP	381.47	172.52	373.92	113.59
Sole parent pension	672.86	225.02	657.74	170.55
Wife/carer pension	363.64	151.26	389.68	178.78
Sickness allowance	356.97	158.91	336.60	100.30
Widow Allowance	389.22	85.56	401.58	82.76
Special benefit	363.57	210.63	323.91	105.96
Partner allowance	337.67	136.44	319.51	149.10
Youth training allowance	200.85	182.47	168.12	79.55
1995/96				
All benefits	380.24	200.91	378.47	156.02
NSA/JSA	352.86	185.40	331.48	159.72
Age Pension	353.20	141.59	375.28	132.32
MAA	386.76	213.11	337.76	101.13
DSP	367.13	226.24	357.87	110.23
Sole parent pension	641.70	224.30	584.23	164.75
Wife/carer pension	387.06	330.21	367.79	169.87
Sickness allowance	371.11	282.50	330.56	86.96
Widow Allowance	375.10	149.53	390.70	111.29
Special benefit	380.16	189.49	309.09	98.69
Partner allowance	385.48	206.84	307.93	130.81
Youth training allowance	177.21	54.77	154.99	81.88
1994/95				
All benefits	371.59	198.22	361.11	138.12
NSA/JSA	375.74	241.93	325.25	145.96
Age Pension	344.69	131.16	368.20	126.11
MAA	358.64	190.84	323.17	85.85
DSP	335.53	176.96	344.75	108.28
Sole parent pension	617.19	271.63	504.32	137.71
Wife/carer pension	331.39	184.46	345.51	154.73
Sickness allowance	350.99	265.95	309.92	86.45
Widow Allowance	375.54	121.32	378.99	117.60
Special benefit	347.97	173.68	293.71	97.88
Partner allowance	388.31	193.34	311.32	114.97
Youth training allowance	331.78	304.62	145.75	85.64

Appendix 5: Comparison of cumulative distributions of annual benefit income by payment type

Figure A1: Annual income from Newstart Allowance, 1996/97

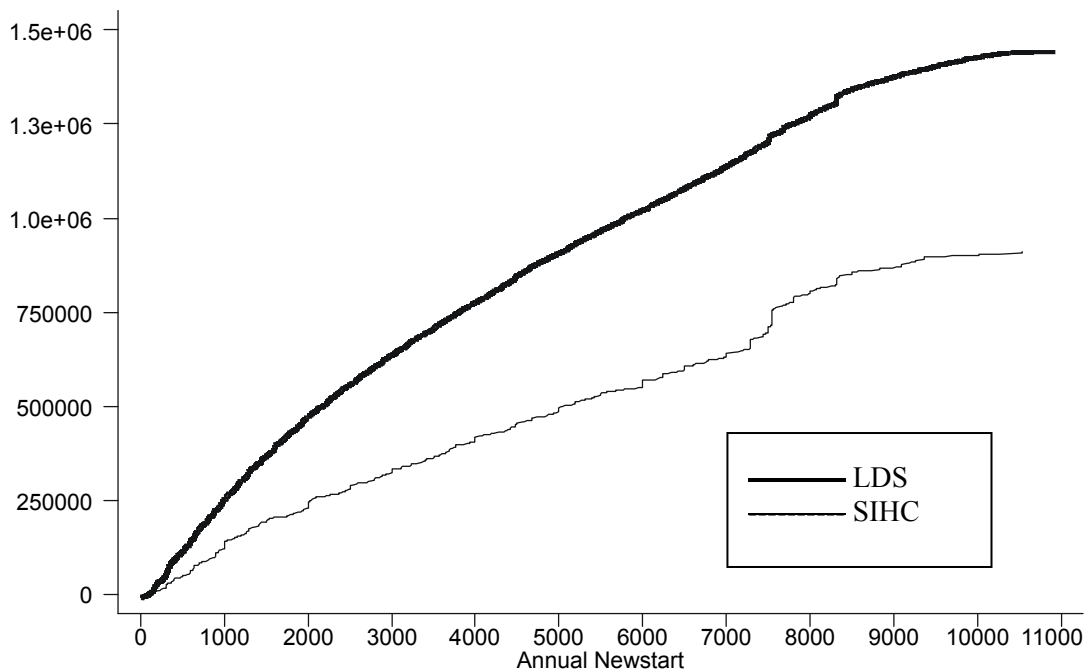


Figure A2: Annual income from Youth Training Allowance, 1996/97

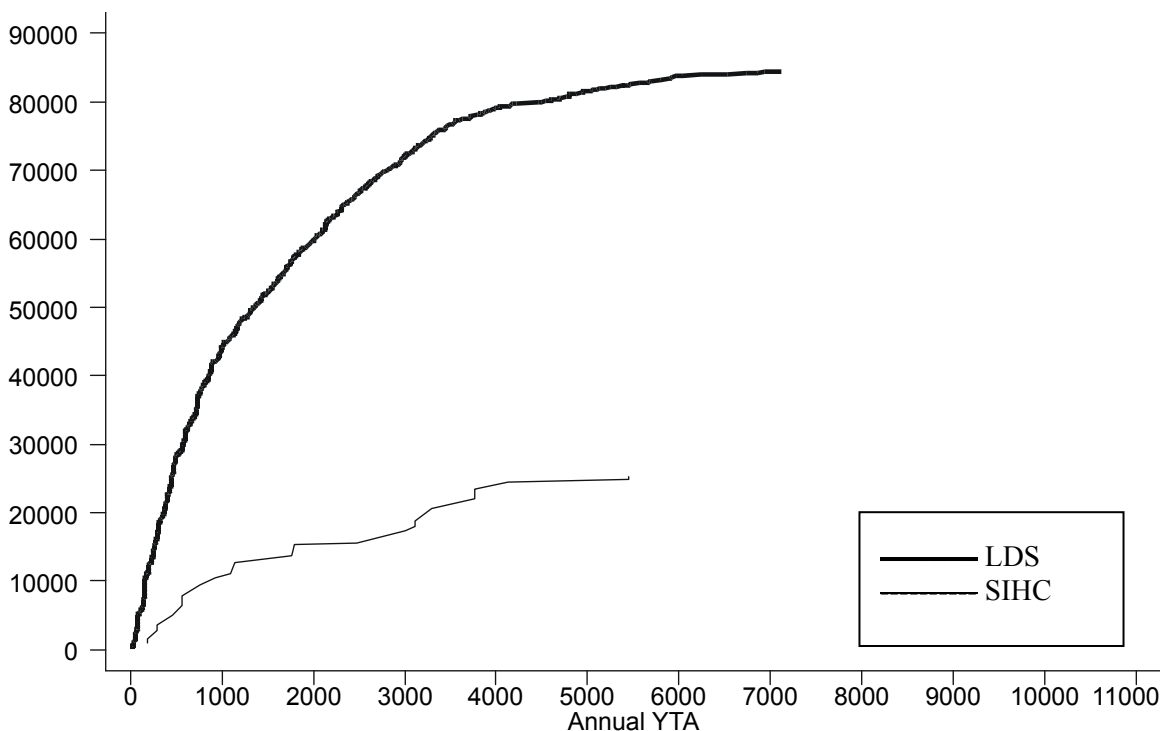


Figure A3: Annual income from Age Pension, 1996/97

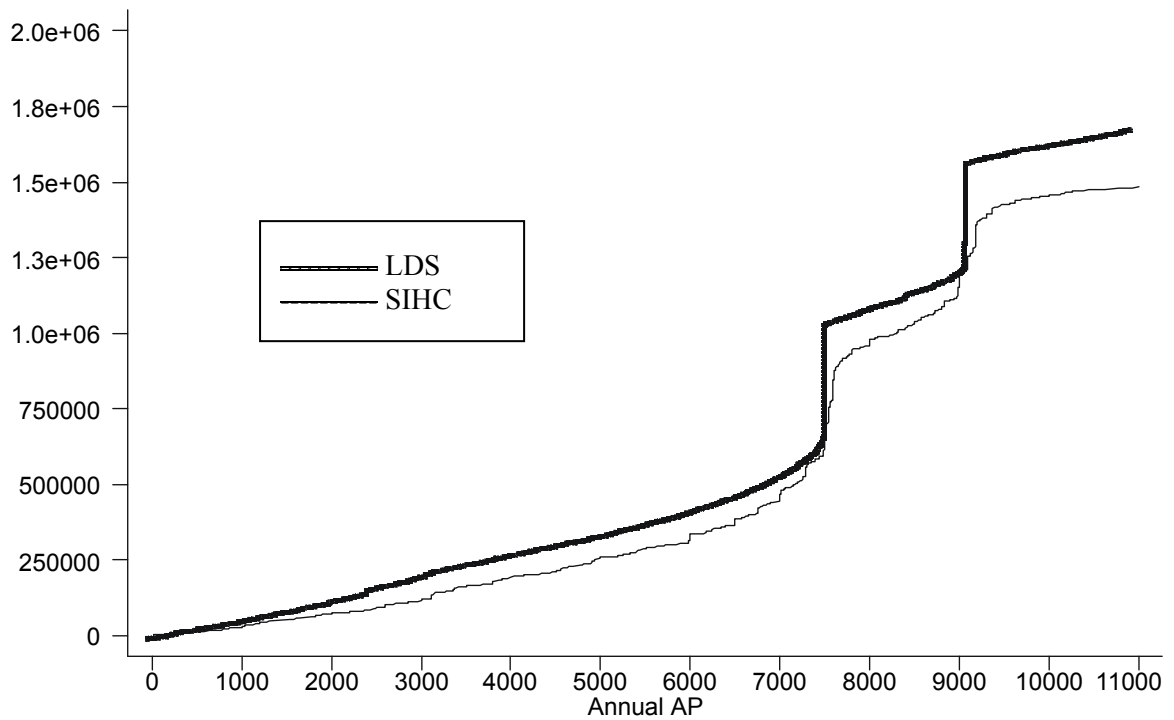


Figure A4: Annual income from Disability Support Pension, 1996/97

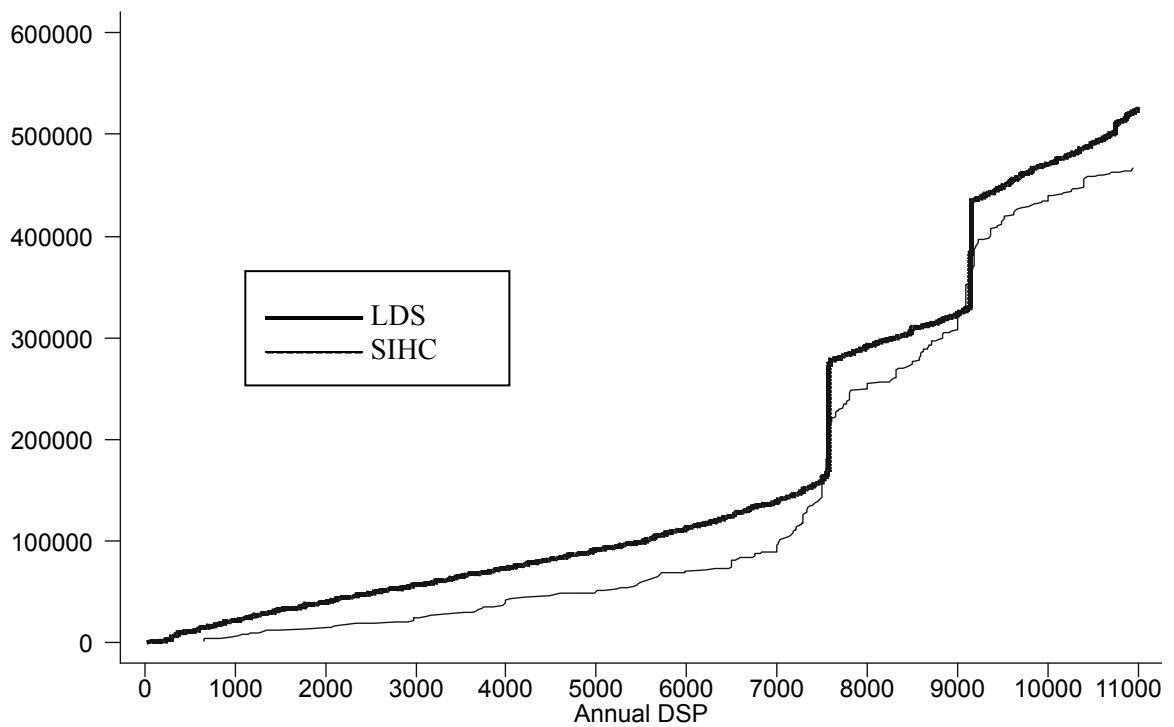


Figure A5: Annual income from Sole Parent Pension, 1996/97

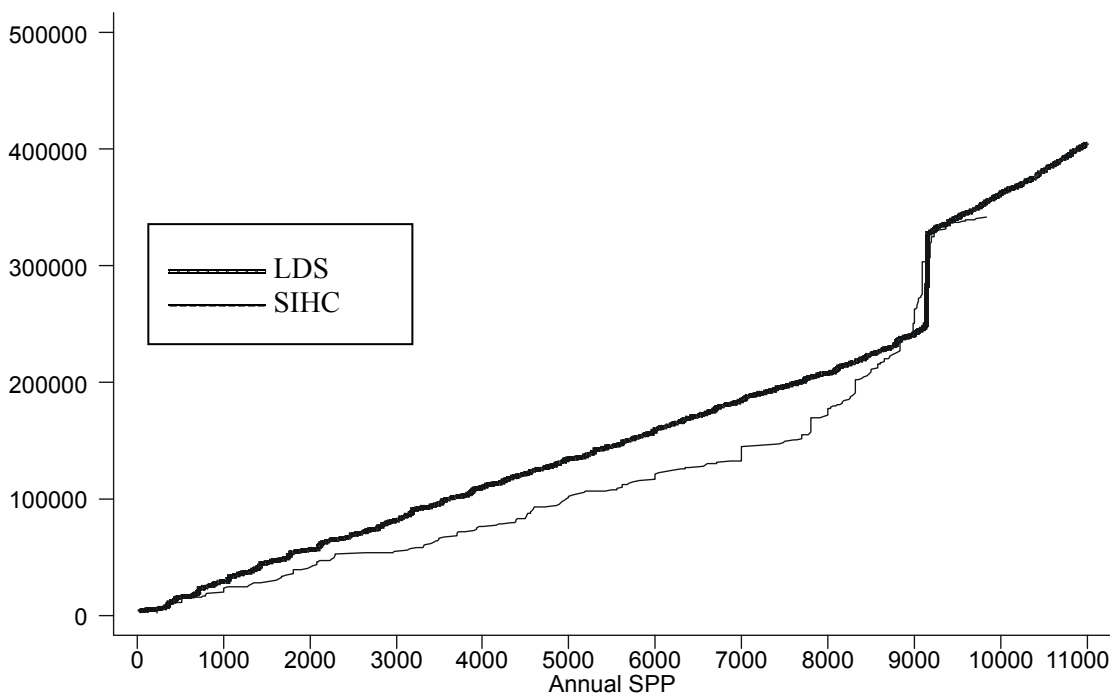


Figure A6: Annual income from Sickness Allowance, 1996/97

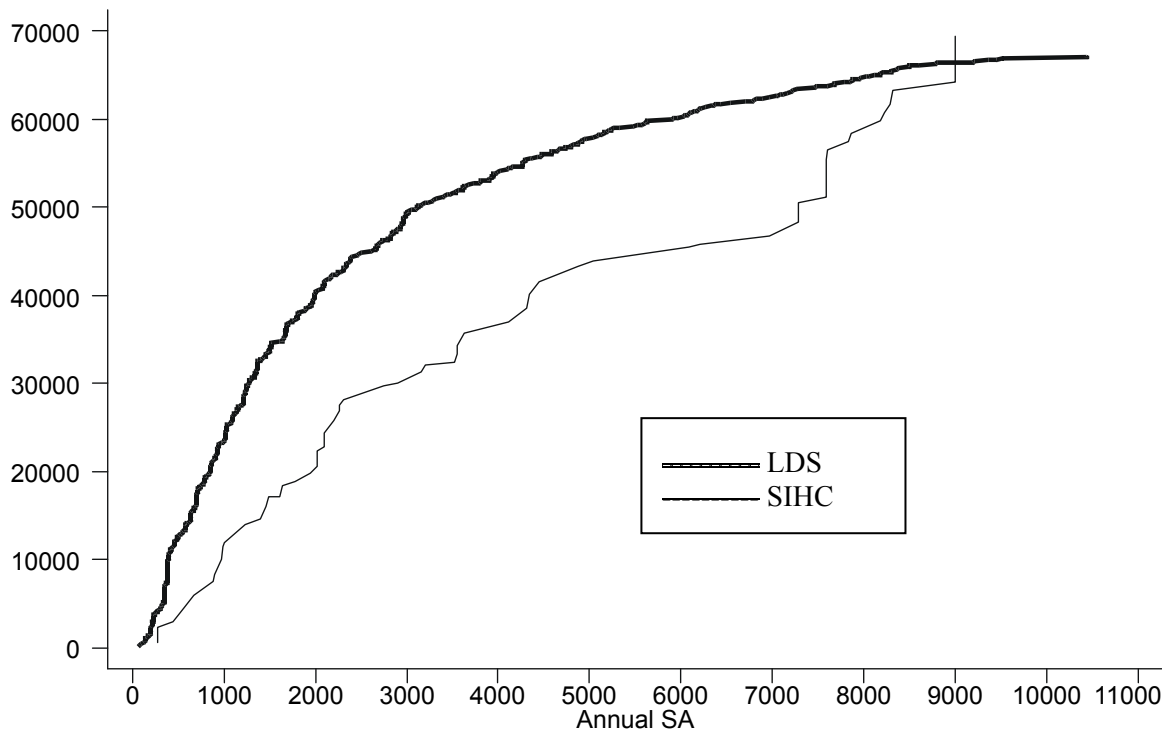


Figure A7: Annual income from Partner Allowance, 1996/97

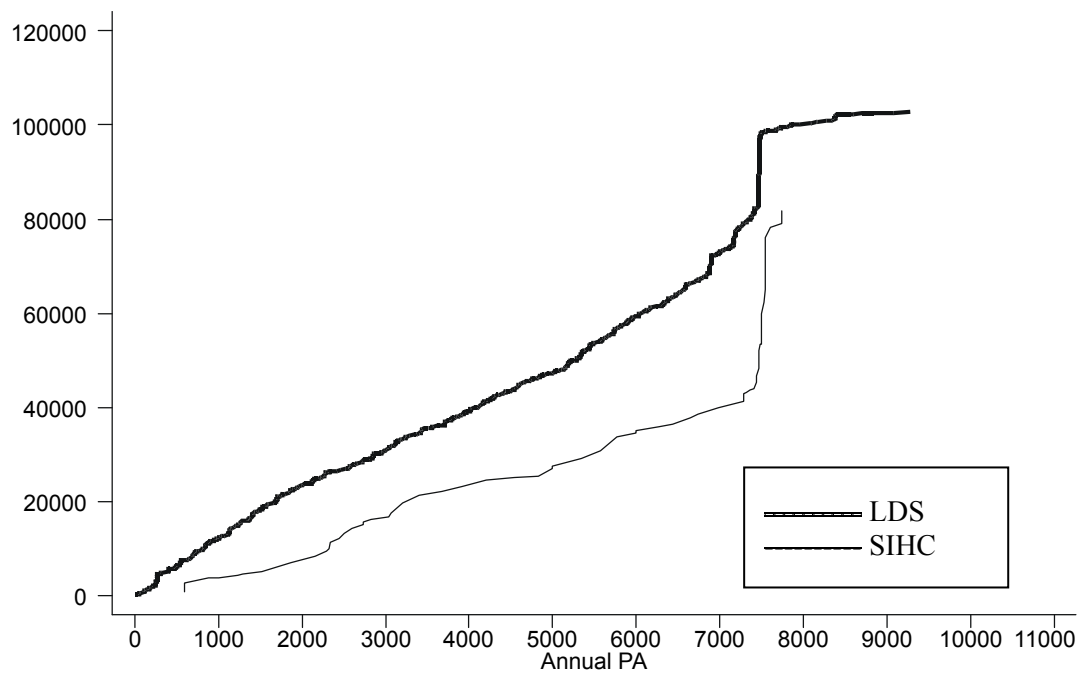
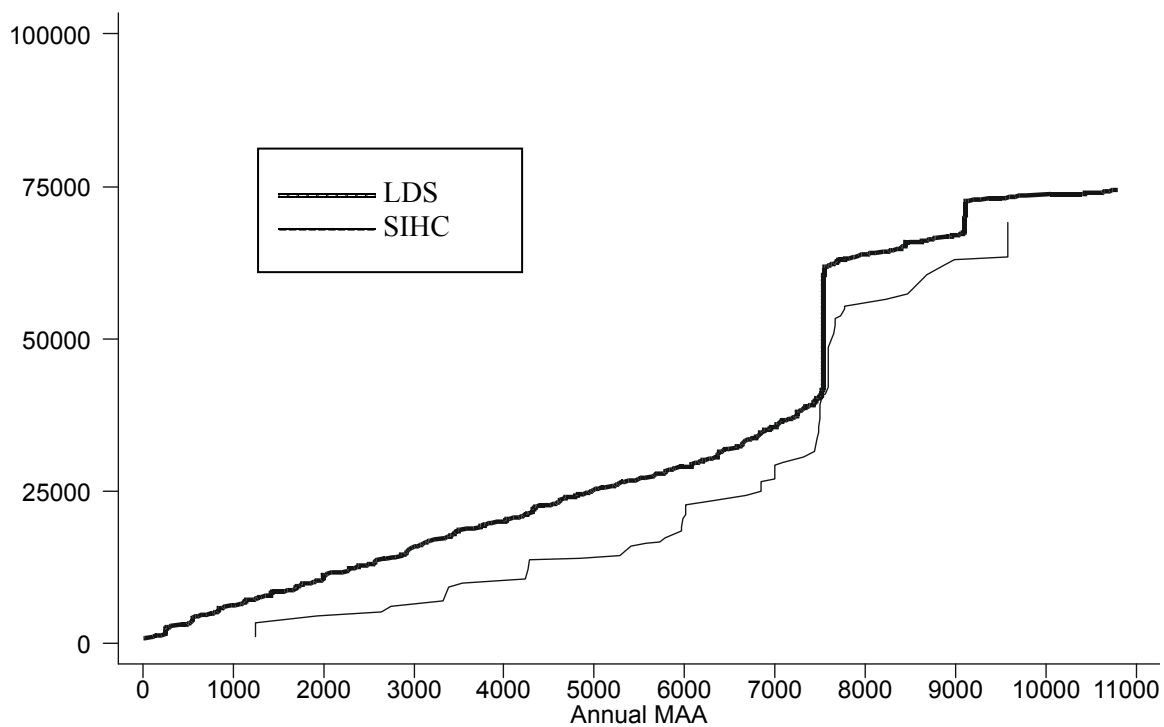


Figure A8: Annual income from Mature Age Allowance, 1996/97



Appendix 6: Fortnightly benefit income by payment type

Figure A9: Fortnightly income from Newstart Allowance, 1997/98

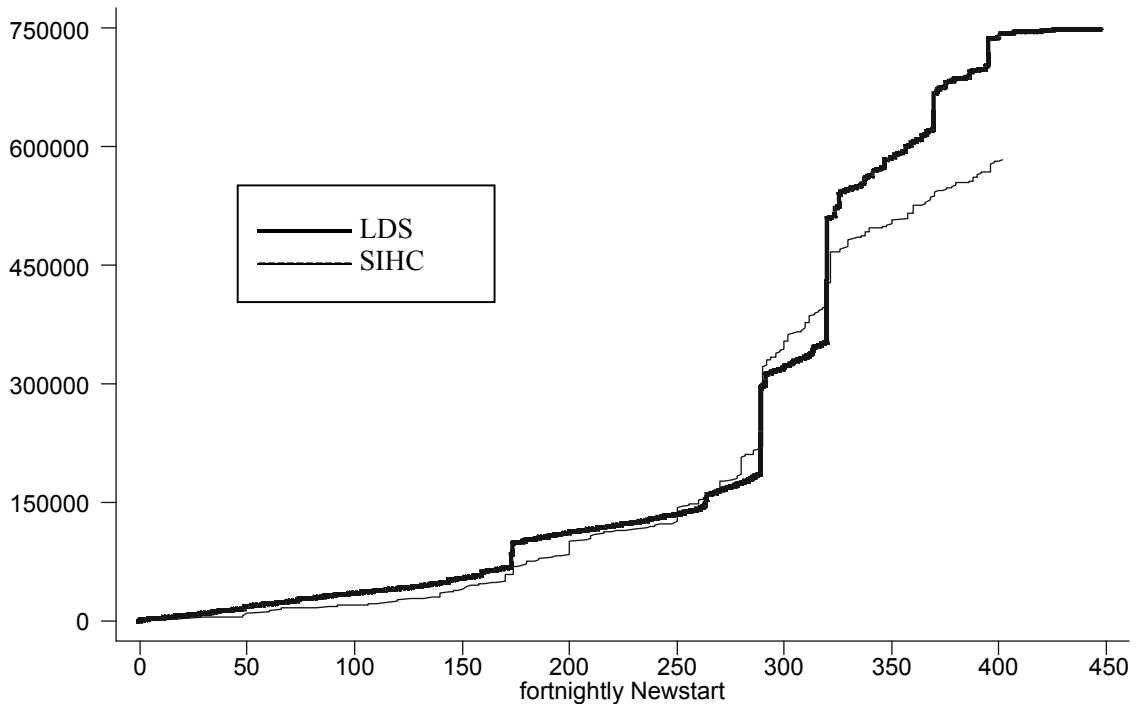


Figure A10: Fortnightly income from YTA, 1997/98

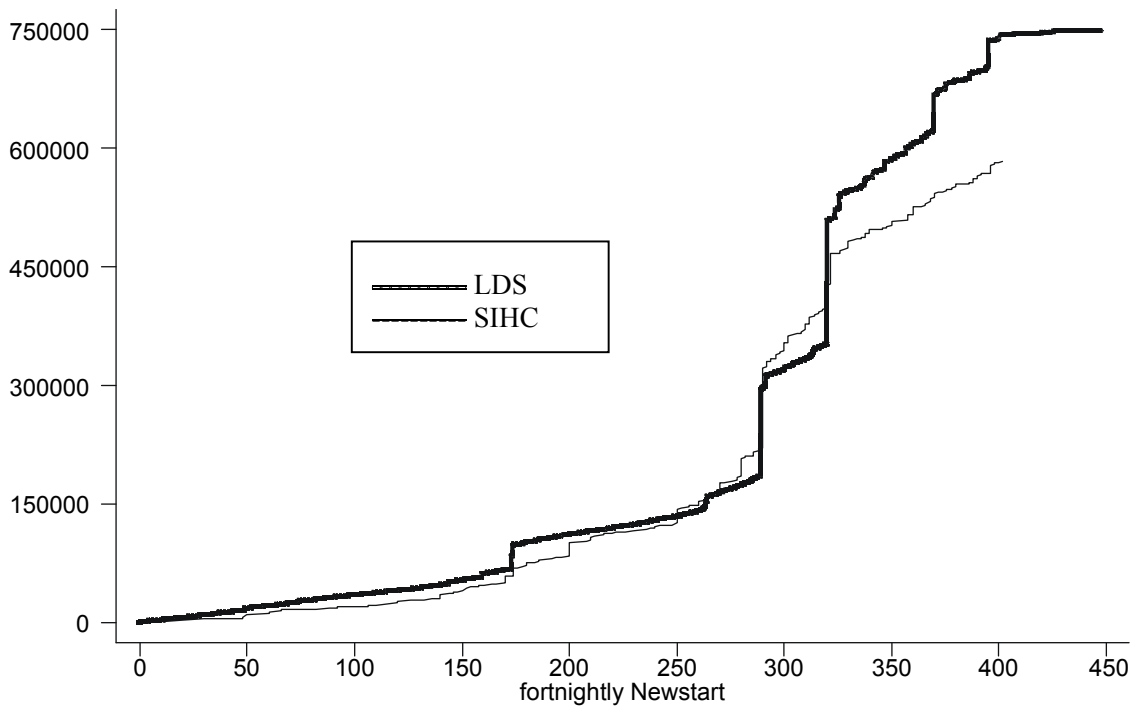


Figure A11: Fortnightly income from AP, 1997/98

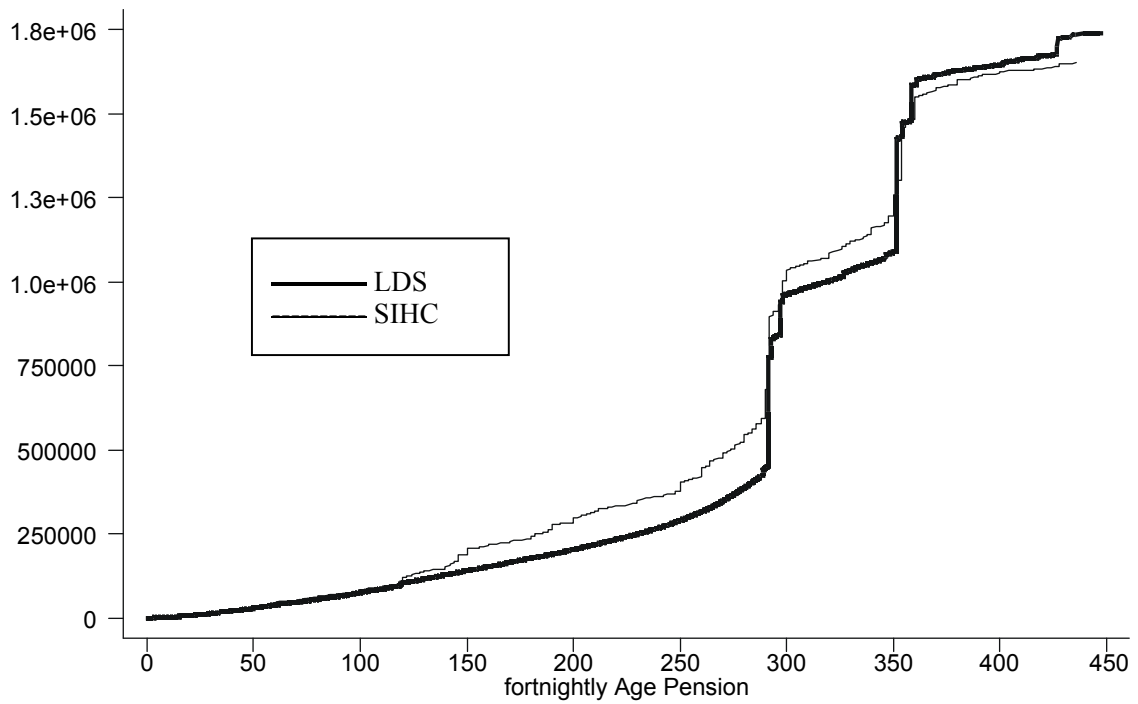


Figure A12: Fortnightly income from Disability Support Pension, 1997/98

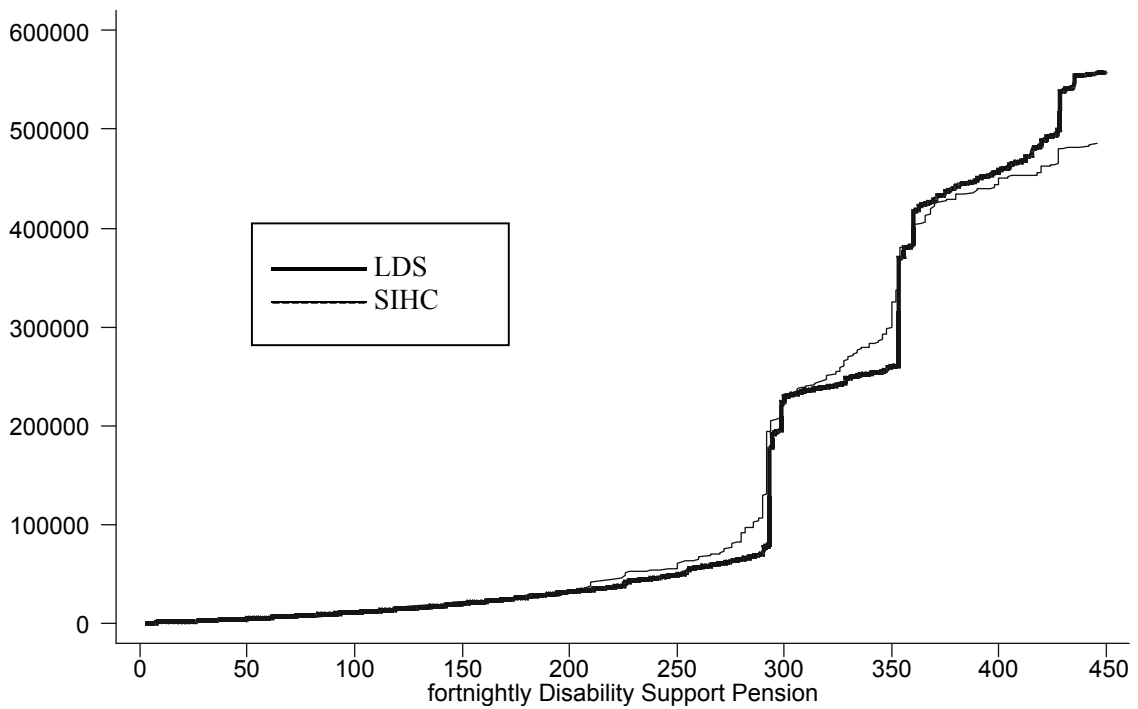


Figure A13: Fortnightly income from Sole Parent Pension, 1997/98

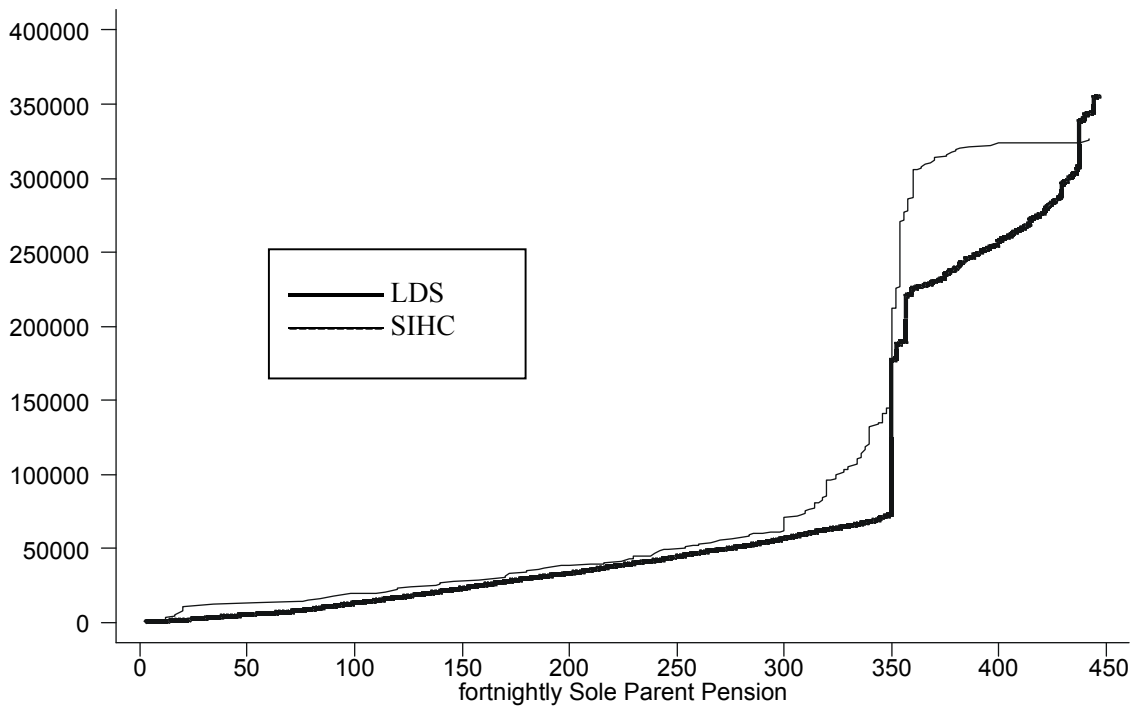


Figure A14: Fortnightly income from Sickness Allowance, 1997/98

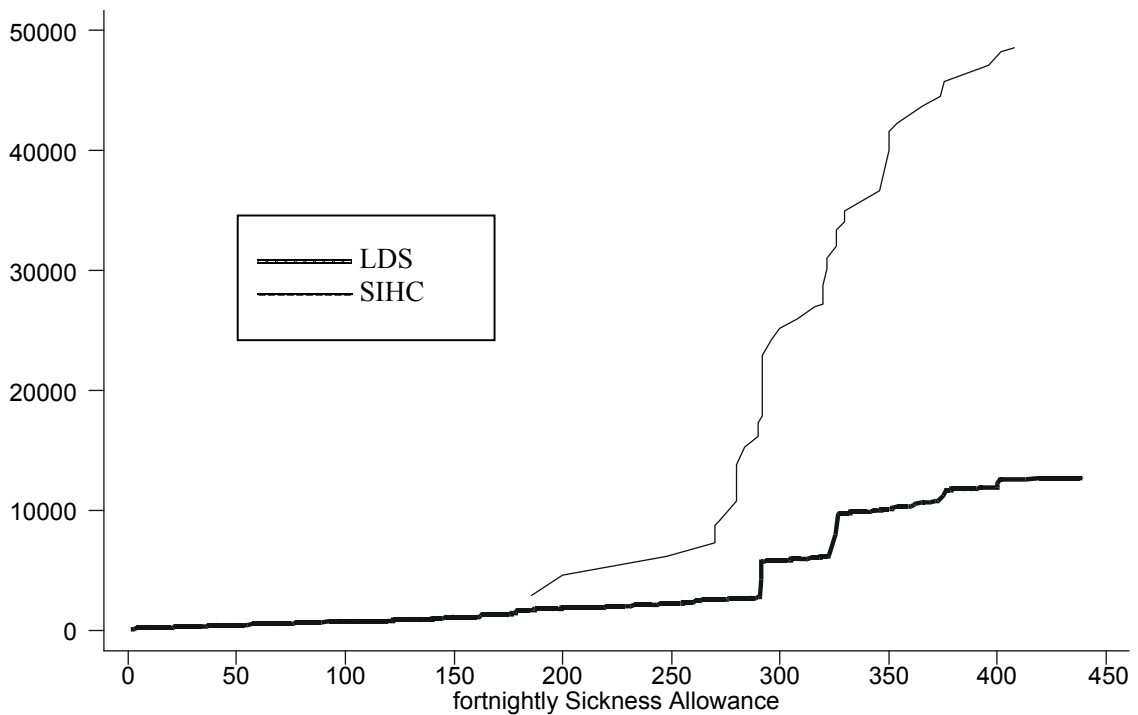


Figure A15: Fortnightly income from Partner Allowance, 1997/98

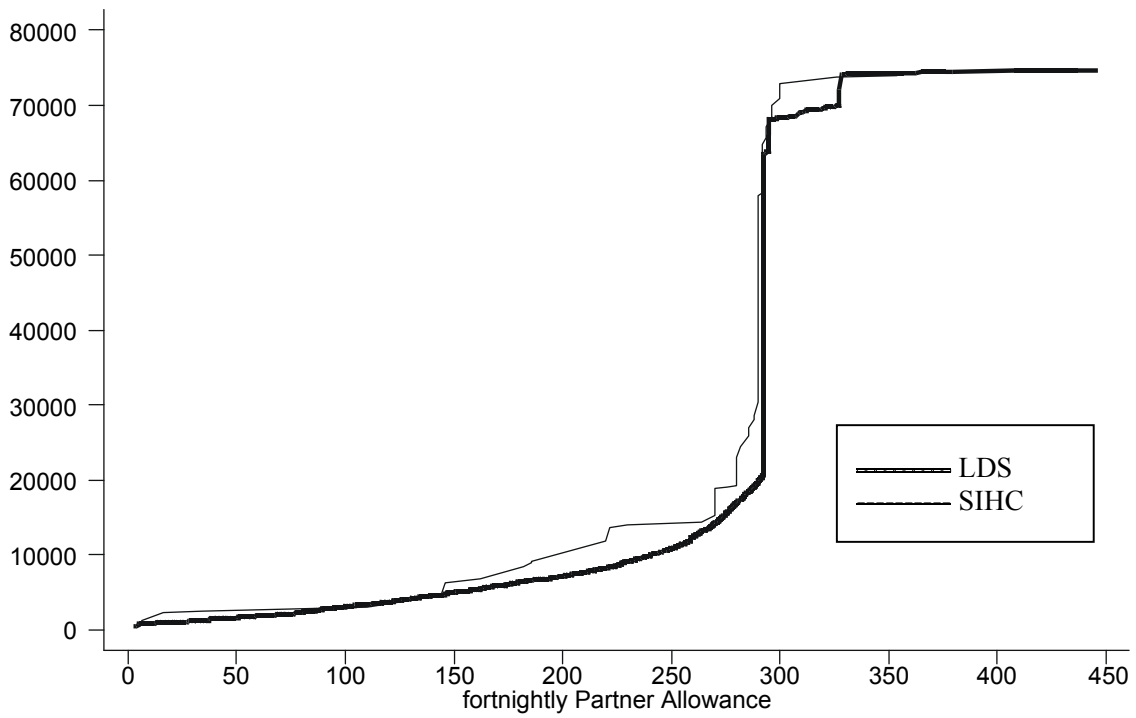
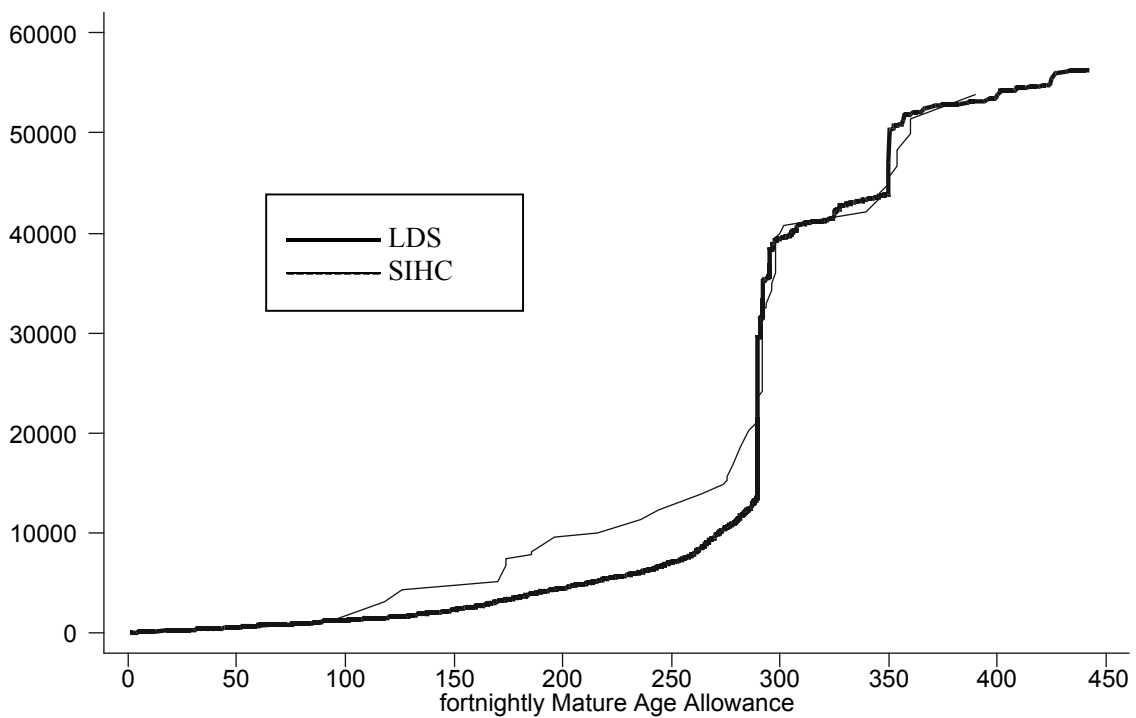


Figure A16: Fortnightly income from Mature Age Allowance, 1997/98



Appendix 7: Difference in total fortnightly incomes by payment type

Figure A17: Total fortnightly income for Newstart Allowance recipients, 1997/98

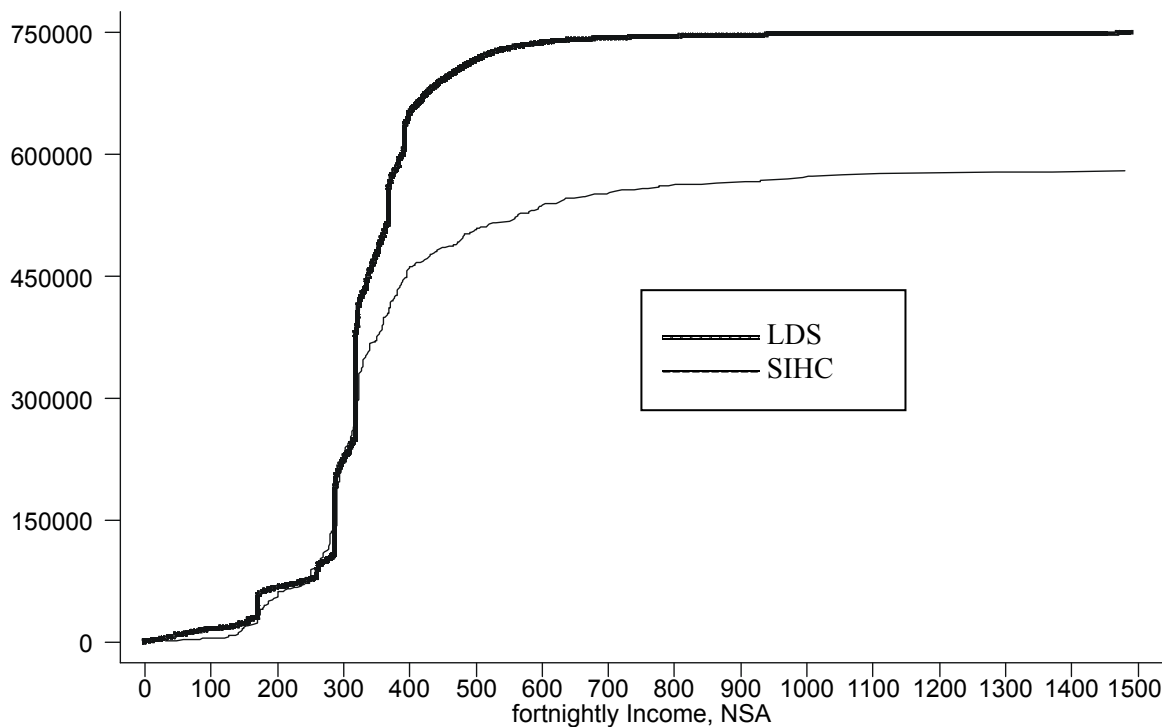


Figure A18: Total fortnightly income for Youth Training Allowance recipients, 1997/98

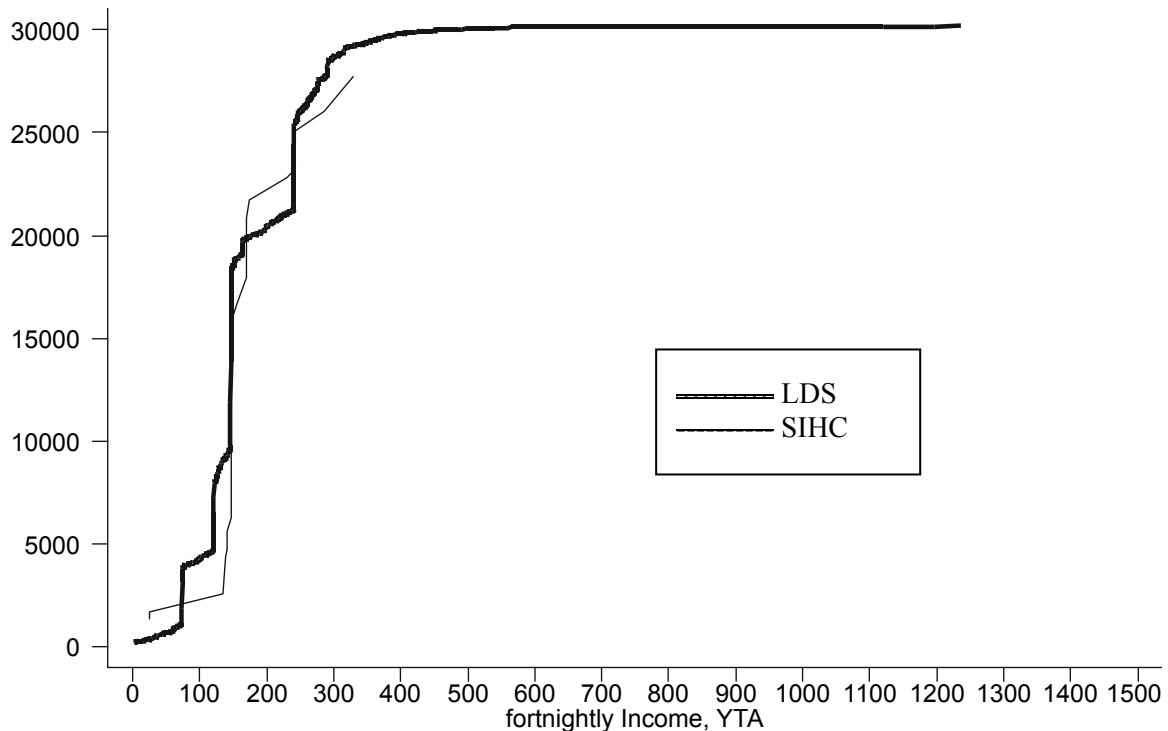


Figure A19: Total fortnightly income for Age Pensioners, 1997/98

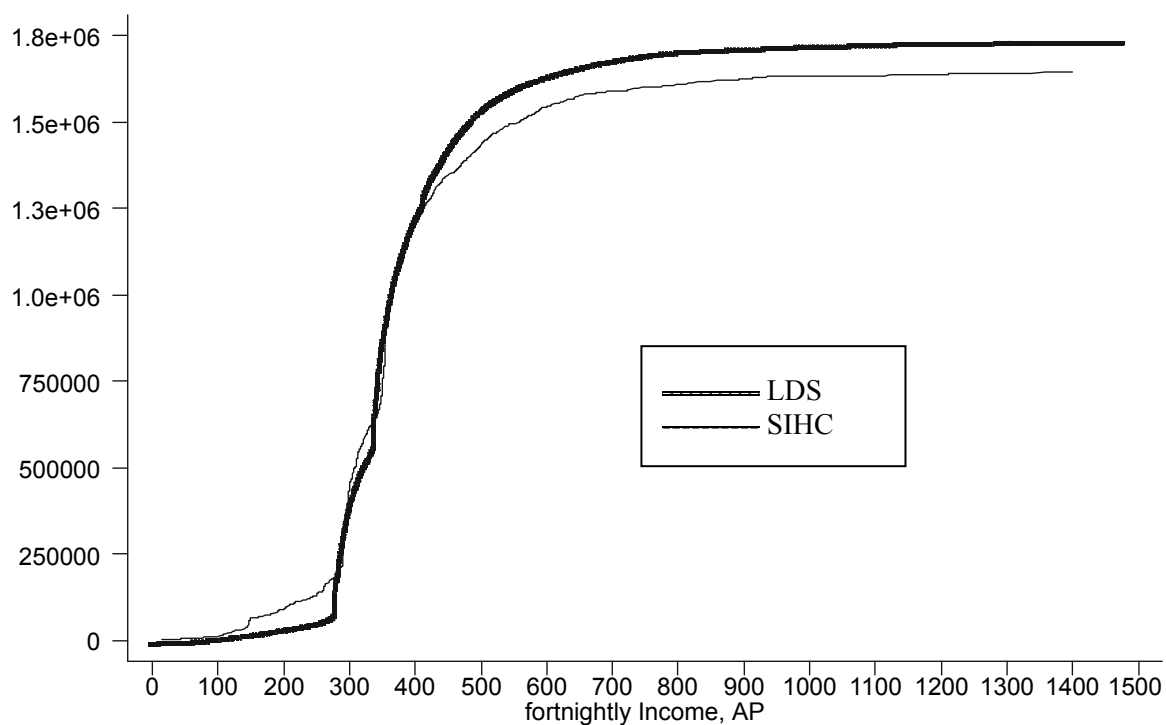


Figure A20: Total fortnightly income for Disability Support Pensioners, 1997/98

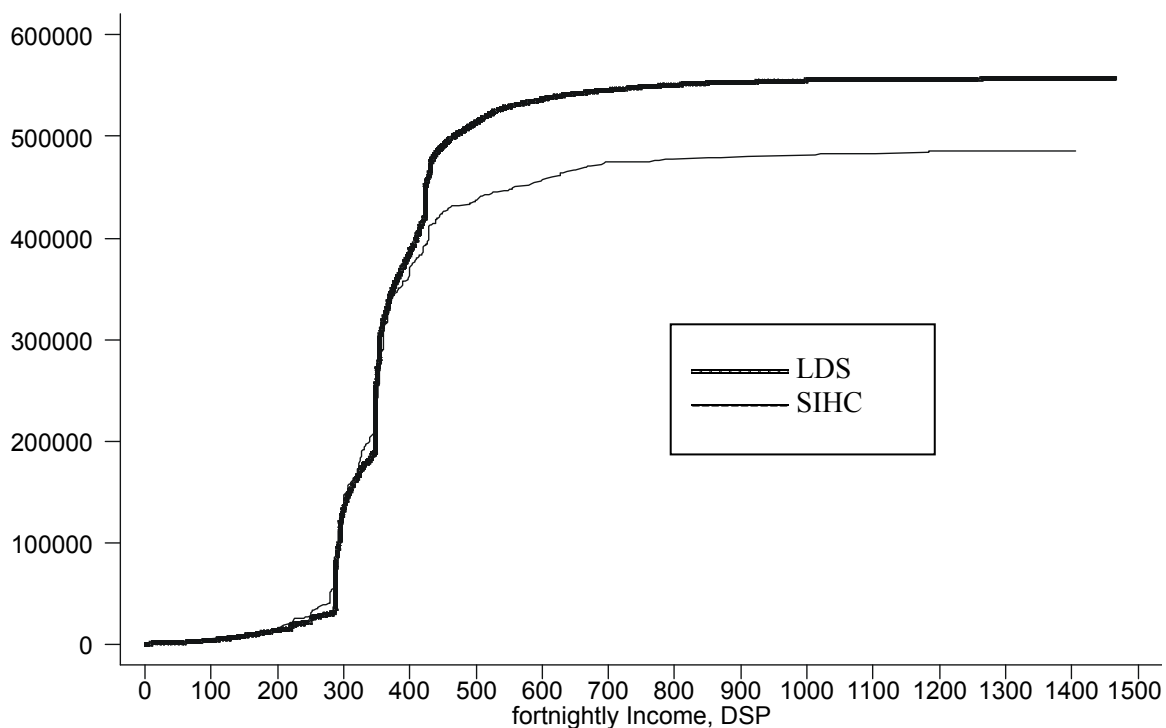


Figure A21: Total fortnightly income for Sole Parent Pensioners, 1997/98

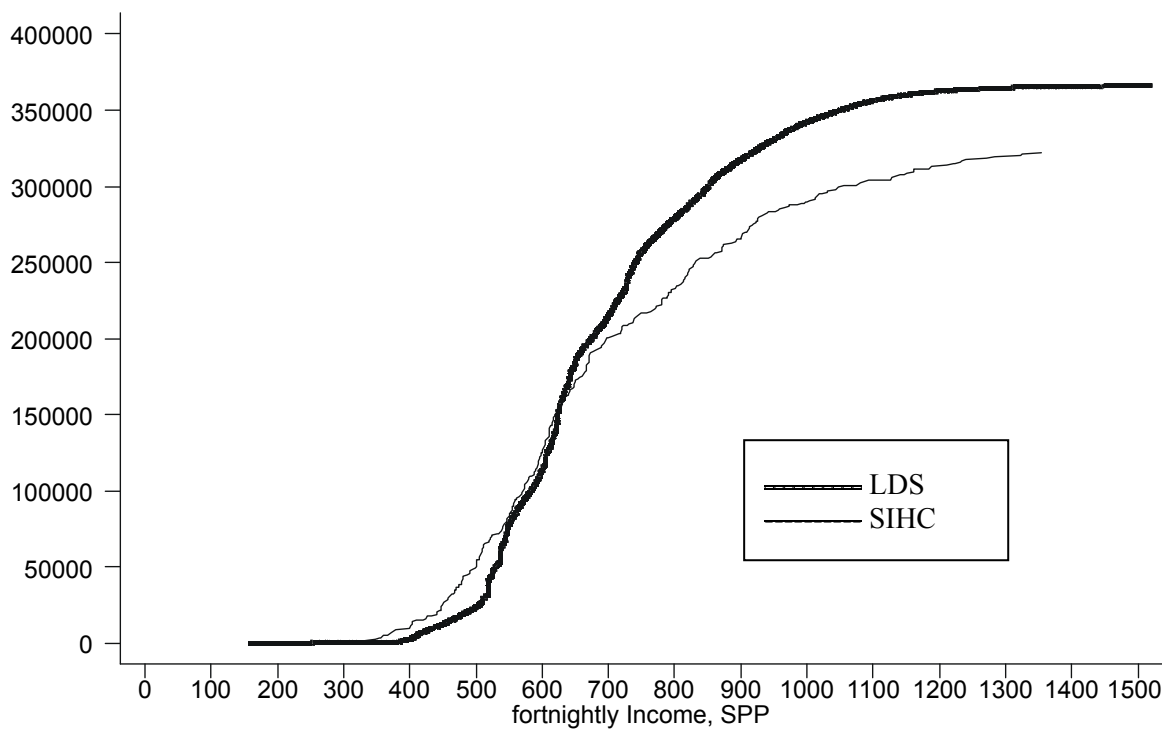


Figure A22: Total fortnightly income for Sickness Allowance recipients, 1997/98

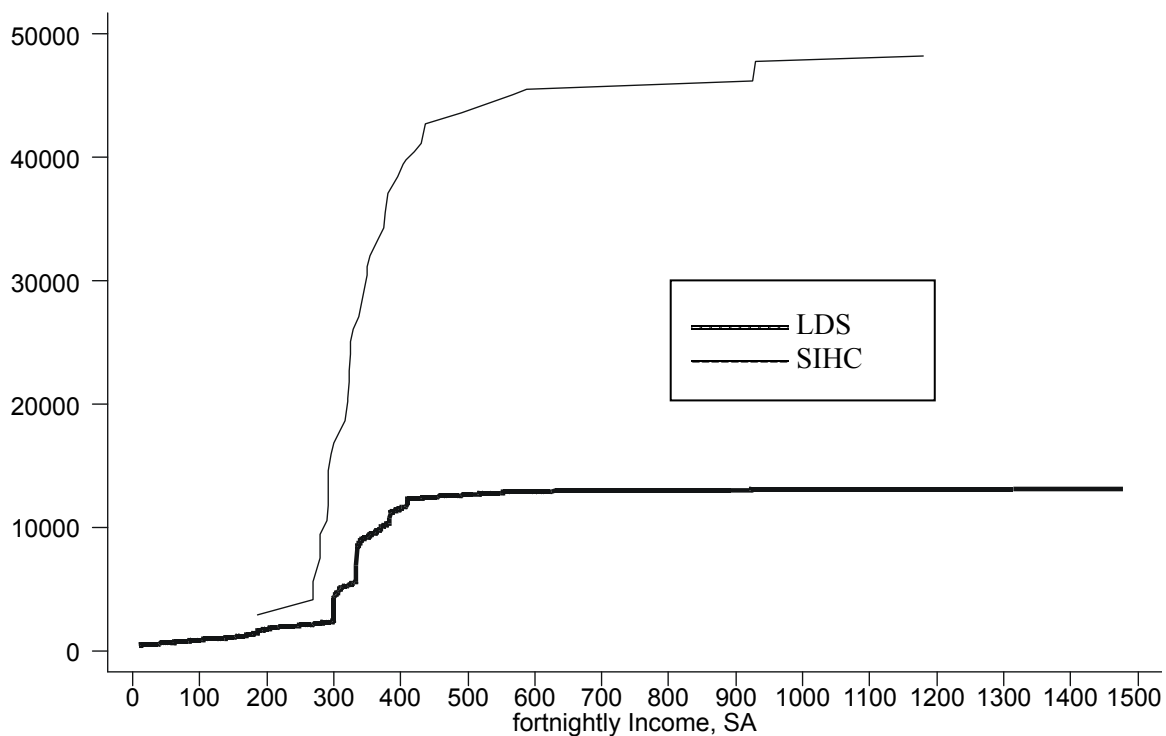


Figure A23: Total fortnightly income for Partner Allowance recipients, 1997/98

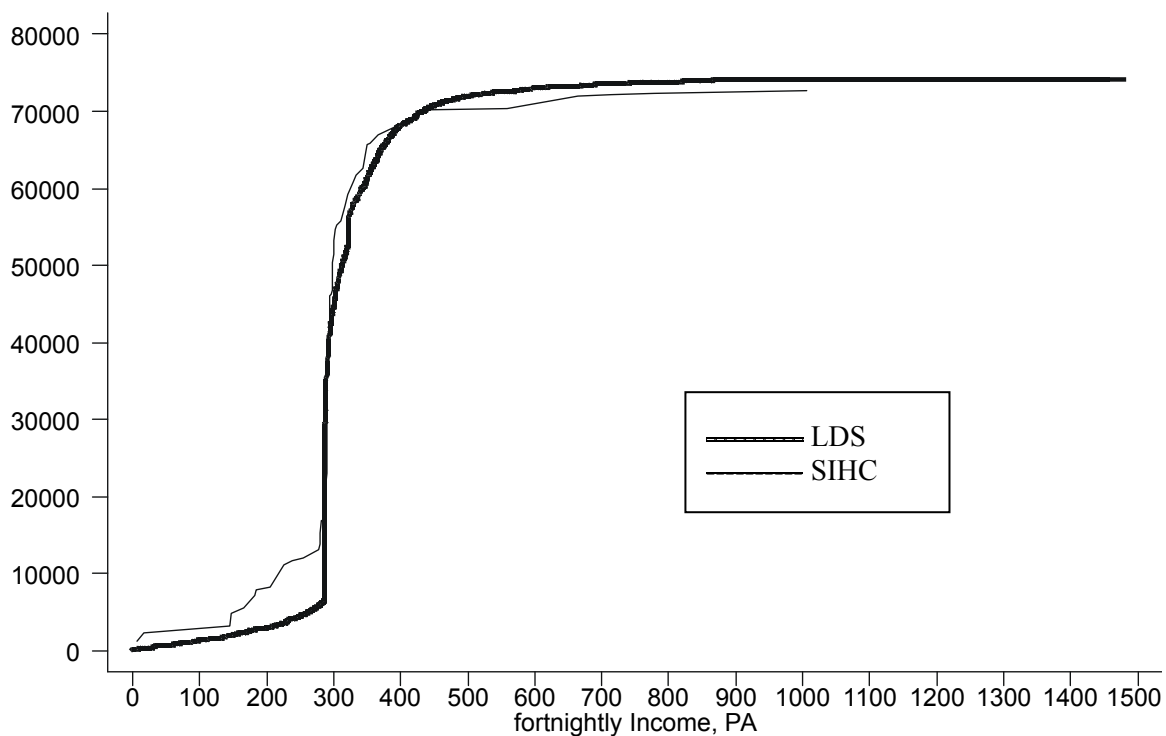
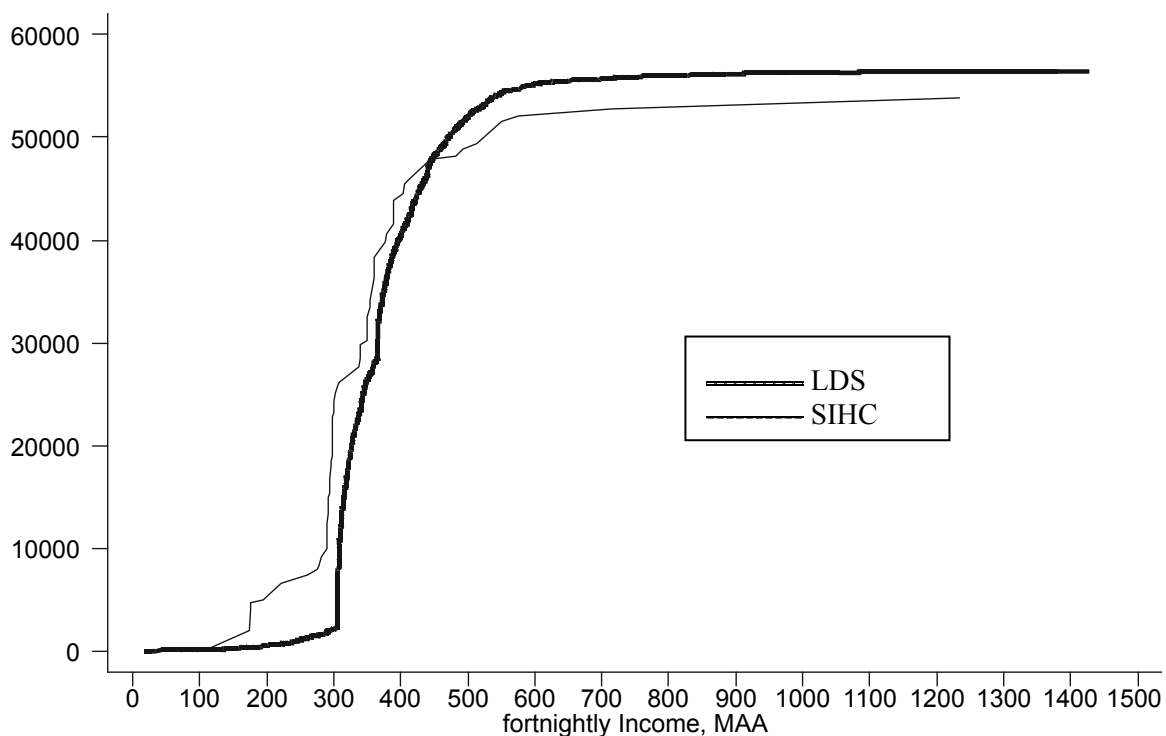


Figure A24: Total fortnightly income for Mature Age Allowance recipients, 1997/98



Appendix 8: An approach to poverty measurement

Estimating the extent and change in poverty is a matter of great public interest. However frequently poverty calculations are made and publicised in a manner that does not admit the limitations and qualifications that are necessarily attached to them. A very useful research project would be one in which the limitations and qualifications associated with poverty measurement were explicitly explored. I outline some matters that would need to be addressed in such a project. The project involves many choices that will influence the results just as much as data limitations. Choices that need to be made include:

1. Choice of uprating poverty line over time. Depending on purpose;
 - an explicit choice of a measure that investigates the purchasing power of the poverty line (eg. CPI).
 - explicit choice of a relative measure that maintains a constant relationship between the poverty line and mean income eg. household disposable income per head. [This becomes an inequality index with a focus on the poor and would be similar to an index comparing low/mean incomes over time]
2. To provide for different views about the selection of the poverty line calculate current poverty rates with different reference levels
3. Explicit choice of equivalence scale. There are many from which to choose, but in the absence of compelling evidence, simple is best (eg. 1/.6/.3). To explore sensitivity to choice, use different equivalence scales:
4. To explore sensitivity to choice of notion of income, vary the notion and calculate poverty.

Comparisons over time will avoid many problems associated with the selection of the reference level of poverty line. However they will be very sensitive to the method by which the poverty line is uprated. In relation to poverty, using both relative and associated methods are valid, depending on the purpose. It is better to be explicit and transparent. Therefore I favour explicit choice of a reference poverty line. In most circumstances for evaluating government policy over time, uprating with a CPI index will be best; however measuring the poor relative to mean income – uprate with HDI per head. A comprehensive project would report both. For setting poverty lines for comparison with current incomes (social security benefits, wages etc) uprate poverty line with an income based index (HDI per head).