Imputation of Pension Accruals and Investment Income in Survey Data

Andrew Aitken[1,2] and Martin Weale[3,4]

[1,3]Economic Statistics Centre of Excellence

[2]National Institute of Economic and Social Research

[4]Centre for Macroeconomics and King's College, London

ESCoE Discussion Paper 2018-05

March 2018

ISSN 2515-4664

About the Economic Statistics Centre of Excellence (ESCoE)

The Economic Statistics Centre of Excellence provides research that addresses the challenges of measuring the modern economy, as recommended by Professor Sir Charles Bean in his Independent Review of UK Economics Statistics. ESCoE is an independent research centre sponsored by the Office for National Statistics (ONS). Key areas of investigation include: National Accounts and Beyond GDP, Productivity and the Modern economy, Regional and Labour Market statistics.

ESCoE is made up of a consortium of leading institutions led by the National Institute of Economic and Social Research (NIESR) with King's College London, innovation foundation Nesta, University of Cambridge, Warwick Business School (University of Warwick) and Strathclyde Business School.

ESCoE Discussion Papers describe research in progress by the author(s) and are published to elicit comments and to further debate. Any views expressed are solely those of the author(s) and so cannot be taken to represent those of the ESCoE, its partner institutions or the ONS.

For more information on ESCoE see www.escoe.ac.uk.

Contact Details
Economic Statistics Centre of Excellence
National Institute of Economic and Social Research
2 Dean Trench St
London SW1P 3HE
United Kingdom

T: +44 (0)20 7222 7665
E: escoeinfo@niesr.ac.uk

Imputation of Pension Accruals and Investment Income in Survey Data

Andrew Aitken[1,2] and Martin Weale[3,4]

[1,3]Economic Statistics Centre of Excellence, [2]National Institute of Economic and Social Research, [4]Centre for Macroeconomics and King's College, London

Abstract
This paper explores the problem of augmenting the data in the UK's Living Costs and Food Survey in order to address two issues. First we are concerned with broadening the definition of income to include accrual of pension rights and secondly we aim to address the point that investment incomes are materially under-recorded. We draw on the Wealth and Assets Survey to address the first point and the Survey of Personal Incomes for the second. We present an approach to stochastic imputation which largely replicates the distributional properties of the source data and show how it can be adapted to address the issue of covariance between the variables imputed. Our initial results suggest that imputation of pension accruals raises both the Gini coefficient and the geometric mean of equivalised household income materially, while the effects of imputing investment income are more marked on the Gini coefficient than on the geometric mean of household income.

Key words: income distribution, inequality aversion, welfare indicator, cost of living

JEL classification: I31, D12, E21, C83, E20

Contact Details
Martin Weale
Economic Statistics Centre of Excellence
National Institute of Economic and Social Research
2 Dean Trench St
London SW1P 3HE
United Kingdom

Email: a.aitken@niesr.ac.uk, martin.weale@outlook.com

This ESCoE paper was first published in March 2018.

Imputation of Pension Accruals and Investment Income in Survey Data

Andrew Aitken* and Martin Weale†
Economic Statistics Centre of Excellence
National Institute of Economic and Social Research
2, Dean Trench Street,
London SW1P 3HE

27th March 2018

Abstract

This paper explores the problem of augmenting the data in the UK's Living Costs and Food Survey in order to address two issues. First we are concerned with broadening the definition of income to include accrual of pension rights and secondly we aim to address the point that investment incomes are materially under-recorded. We draw on the Wealth and Assets Survey to address the first point and the Survey of Personal Incomes for the second. We present an approach to stochastic imputation which largely replicates the distributional properties of the source data and show how it can be adapted to address the issue of covariance between the variables imputed. Our initial results suggest that imputation of pension accruals raises both the Gini coefficient and the geometric mean of equivalised household income materially, while the effects of imputing investment income are more marked on the Gini coefficient than on the geometric mean of household income.

JEL codes: I31, D12, E21, C83, E20
Keywords: Income distribution, Inequality Aversion, Welfare Indicator, Cost of Living

*National Institute of Economic and Social Research and Economic Statistics Centre of Excellence.
†King's College, London, Economic Statistics Centre of Excellence and Centre for Macroeconomics. martin.weale@outlook.com

1 Introduction

This paper describes some of the building blocks needed in the construction of a democratic measure of national income growth. The limitations of GDP are widely known. It is, in contrast to the original measure used in the national accounts (Meade and Stone 1941) measured gross rather than net of depreciation and it does not take any account of property income due abroad or received from abroad. Thus the Stiglitz Commission (Stiglitz, Sen, and Fitoussi 2009) suggested more focus on income rather than GDP, and it is also natural to focus on measures *per capita* or per household.

Macroeconomic aggregates, whether real GDP or measures of real income however, as indicators of the growth of real income have the effect of weighting together the growth experience of the individual people/households according to their shares in total income. Thus the experience of high-income households contributes more towards the aggregate than does the experience of low- income households. One may therefore describe growth in such measures as plutocratic rather than democratic.

A democratic measure of growth would, in contrast, average the growth experience of every household and would therefore be structured round the growth in the geometric mean of household income.[1] While Fleurbaey (2009) is sceptical about the use of income-based aggregates as indicators of well-being, it is possible to build on the approach of Sefton and Weale (2006) to establish a link between democratic income growth and welfare growth, at least subject to the assumptions that consumption and savings decisions are the result of inter-temporal optimisation. While we do not claim any stronger connection between our proposed measure and welfare, it is worth noting that Layard, Nickell, and Mayraz (2008) find an elasticity between marginal utility and income in the range of 1.19 to 1.34 depending on the country studied. Thus the indicator we are developing is probably not very far from a welfare indicator.

Historically and, we suspect for some time into the future, the information on the distribution of income that we require to construct our democratic measure of national income will be provided by household surveys. Even with increased access to administrative data since the Digital Economy Act (2017) became law, it seems likely that

[1] In producing democratic measures of real income growth issues also arise over the choice of price index. These are not discussed here.

it will remain necessary to rely on these rather than simply draw on tax records. One reason for this is that, in the United Kingdom at least, income tax is charged on an individual basis rather than a household basis. As a consequence, tax data cannot, even if available, obviously be linked to households rather than individuals.[2] On the other hand, since there are thought to be economies of scale associated with household size, distributional statistics are usually calculated after dividing money incomes by equivalence scales which, even for adults, are concave in household size. For this reason the distributional national accounts proposed by Piketty, Saez, and Zucman (2018) do not contain the information required.

Nevertheless there is every reason to feel uncomfortable with the use of survey data. There are often large discrepancies between the survey totals and the equivalent variables shown in aggregate national accounts data as Tonkin (2015) and Brewer, Etheridge, and O'Dea (2017) show for the UK and Fesseau and Matteonetti (2013) demonstrate internationally. The OECD has therefore recently promoted work designed to reconcile survey data with the national accounts. This is obviously a necessary step in the allocation of national income to households. Undistributed income has to be allocated, in some form, with reference to distributed income but, if measures of this are unsatisfactory, that needs to be addressed before the allocation of undistributed income.

This paper sets out some of the issues which arise in the allocation of pension accruals and investment income to households and indicates the effects of this on the distribution of income. Following ONS practice, we work to the Living Costs and Food Survey (Office for National Statistics and Dept for Environment and Affairs 2017). Because the approach proposed is stochastic, the distribution of estimates of income inequality has to be explored by means of repeated simulation. Pension accruals do not, of course, appear in survey data at all, and the shortfalls for investment income in the Living Costs and Food Survey are considerable. In 2013 only seventy-two per cent of Blue Book investment income was reported in the LCFS. Overall, household income as reported in the LCFS was seventy-three per cent of Blue Book income.

Much work on reconciliation is done simply by scaling (e.g. Fesseau and Matteonetti (2013) and Fixler and Johnson (2014)). This can be done at any desired level of disag-

[2]It may eventually be possible to match people by addresses and thus to approximate households with tax records. But even then only very limited information on household circumstances will be available.

gregation. Thus Jorgenson and Schreyer (2015) reconcile consumption figures by scaling expenditure on each of fifteen categories of consumption separates. A major concern with scaling is that any mis-reported zeros will remain as zeros. This is a very material issue with investment income in the Living Costs and Food Survey since there are considerably more zero entries shown for tax payers than are suggested by the Survey of Personal Incomes (HM Revenue and Customs 2016), compiled from tax returns and tax codes. Reporting zero rather than actual receipts may be a material factor behind the under-recording discussed above; if this is the case use of upward scaling will tend to accentuate estimates of inequality.

The main literature on imputation addresses the problem of missing observations in a single data source (e.g. Rubin (1987) and van Buuren (2007)), but the situation we face is one where we assume at least some of the observations are faulty but we do not know which to accept and which to reject. In such circumstances Piketty, Saez, and Zucman (2018) and Saez and Zucman (2016) suggest drawing on other data sources such as other surveys. If imputation is carried out the basis of covariates which are observed in both the source data and in the survey records of interest then there must be a need for a stochastic term to represent the random variation between the relevant covariates and the data of interest in the source data. Neglect of this will not be an issue if only arithmetic averages are of interest. Since, however, one of the purposes of using survey data is to address distributional issues, the treatment of any stochastic terms is in fact likely to be of considerable importance. For pension accruals it is necessary to impute the value of pension rights and then allocate Blue Book pension fund accruals in proportion to these; that aspect of our approach mirrors Piketty, Saez, and Zucman (2018). For investment income it is possible to impute flows directly.

In imputing pension rights we compare the effects of two different approaches to imputation. The first involves the use of Heckman regression to estimate the probability that any household has non-zero private pension rights and conditionally on this, what its pension rights are. Application of this method assumes that the component of either investment income or pension resources not explained by covariates is log-normally distributed. We suggest that, because the logarithms of the variables of interest, investment income and pension resources are not in fact normally distributed, this is not a very satisfactory way of carrying out the imputation. The imposition of a normal distribution on a variable which in fact has fat tails results in too much, rather than too little probability mass being found in the tails. Thus imputation structured round log-normality can over-state the probability of both very low and very high values of the imputed variable.

Our second approach is designed to avoid this problem. It involves categorising the variable in question (investment income/pension rights) into a discrete number of size bands. We can then use an ordered probit model to explain the probability of each variable being in any particular category and can use the parameters of this to impute values for the relevant variables to the Living Costs and Food Survey. We show that this approach results in distributions which match the distributions of the variables in question in the source data much more closely. With multiple stochastic imputation, the correlation between the different stochastic terms is likely to be an important influence on the overall pattern of income inequality which results.

A particular issue which we need to address in the imputation of multiple variables is the covariation in their stochastic components. Uncorrelated stochastic components added in to income might, depending on their relationship with observed measures of income, have the effect of reducing inequality, while if the components are strongly correlated, inequality might be increased. Drawing on multiple sources for our imputation models limits our ability to produce multivariate models which would represent these covariances. The practical solution we adopt, therefore, is to draw on covariances inferred from the Wealth and Assets Survey, and use these even when other sources provide a more satisfactory basis for imputation.

In the next section we discuss further the allocation of national income to households. This is followed by an account of some of the data issues we face. The next section discusses the data. We follow this with an account of the models used to impute pension rights and investment income, and a discussion of the issue of covariances between the imputed variables. The penultimate section presents the results of some stochastic simulations of estimates of the Gini coefficient and the geometric mean of equivalised household income, and the final section concludes.

2 Allocation of the Whole of National Income to Households

Piketty, Saez, and Zucman (2018) introduce the concept of distributional national accounts. They cast, for the United States, a set of national accounts in which the whole of national income is allocated to households. This is a necessary step in producing the background data from which to produce a fully satisfactory democratic measure of income growth. Their work is structured round tax records, a data source which is not available in the United Kingdom and which, in any case, provides information on individuals rather than households. Most work (see for example Lewbel (1997), Schafer (1999) and Crossley and Pendakur (2010)) suggests that it is better to work with households, adjusted for household size using an equivalence scale, as the fundamental unit for welfare analysis.

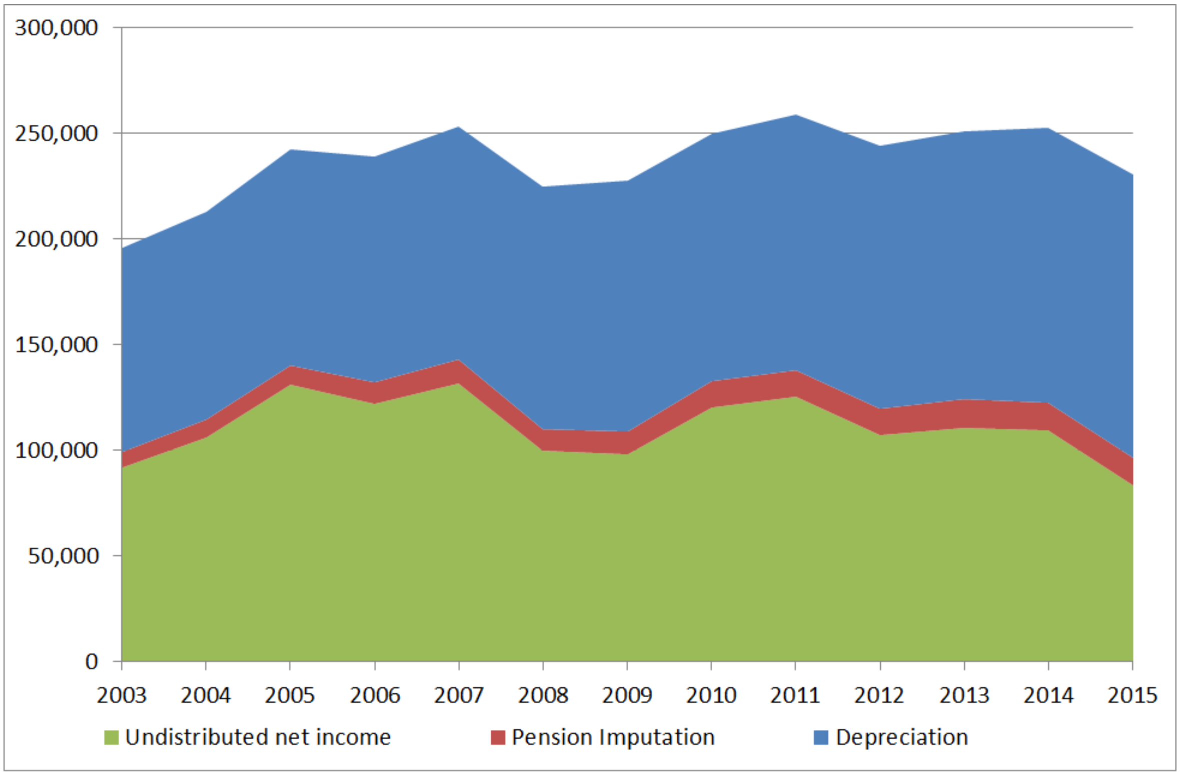
For many years, however, the Office for National Statistics and its predecessor, the Central Statistical Office have used the LCFS to provide estimates of the distribution of household income in the United Kingdom (see Stone and Stone (1972) for an early reference to this work), again after adjusting for household size. These estimates show the distribution of income both before and after the redistributive activities of the government. The latter involve the collection of taxes used to finance the provision of benefits and publicly-provided consumption. Benefits and direct taxes are reported in the survey and indirect taxes are allocated on the basis of reported consumption patterns. The large individual items of public consumption are education and health services (social transfers in kind); these are allocated to the individual households on the basis of what is known about the use of the education system and the use of the health service[3]. No allocation is made of collective consumption, such as public administration and defence. Additionally no allocation is made of the undistributed income of companies or pension fund accruals.

While this ONS work on the LCFS provides the core of an allocation of national income to households there are a number of further steps which are needed. First of all, in common with many household surveys, the aggregates implied are well below those

[3]The allocation of health services is based on average cohort use rather than individual use. It is not assumed that someone who has been provided with expensive health treatment has been a large consumer. Rather it is as though insurance were provided and the insurance is imputed.

shown in the national accounts. Secondly, even when that has been addressed, some basis is needed for allocating the undistributed income of companies and pension funds. Figure 1 shows this undistributed income since 2003. It is, of course, possible to allocate undistributed corporate income on the basis of dividend income but given the issues we identify with the dividend data we do not attempt that at this stage.

Figure 1: The Decomposition of Corporate Gross Disposable Income (£mn)



Source: ONS data, July 2017.

We explain in section 3.2 how we handle pension funds in a manner consistent with the current System of National Accounts and this points to a straightforward treatment of the pension fund transactions which do not directly impinge on households. Imputation is also needed to address collective public consumption and public saving, but that, again, is not pursued further here.

3 Data Issues

We begin this section by indicating the scale of mis-recording. We then consider the issues raised by the treatment of pension funds and pension accruals. This is followed by discussion of the data on pension rights and on investment income, and on the treatment of taxation with imputed income data.

3.1 The Scale of Mis-recording and the Concept of Investment Income

Recent work by Tonkin (2015) provides an indication of the nature of the under-recording in the LCFS relative to the national accounts. This is shown in table 1.

Table 1: Under-recording in the Living Costs and Food Survey, 2013

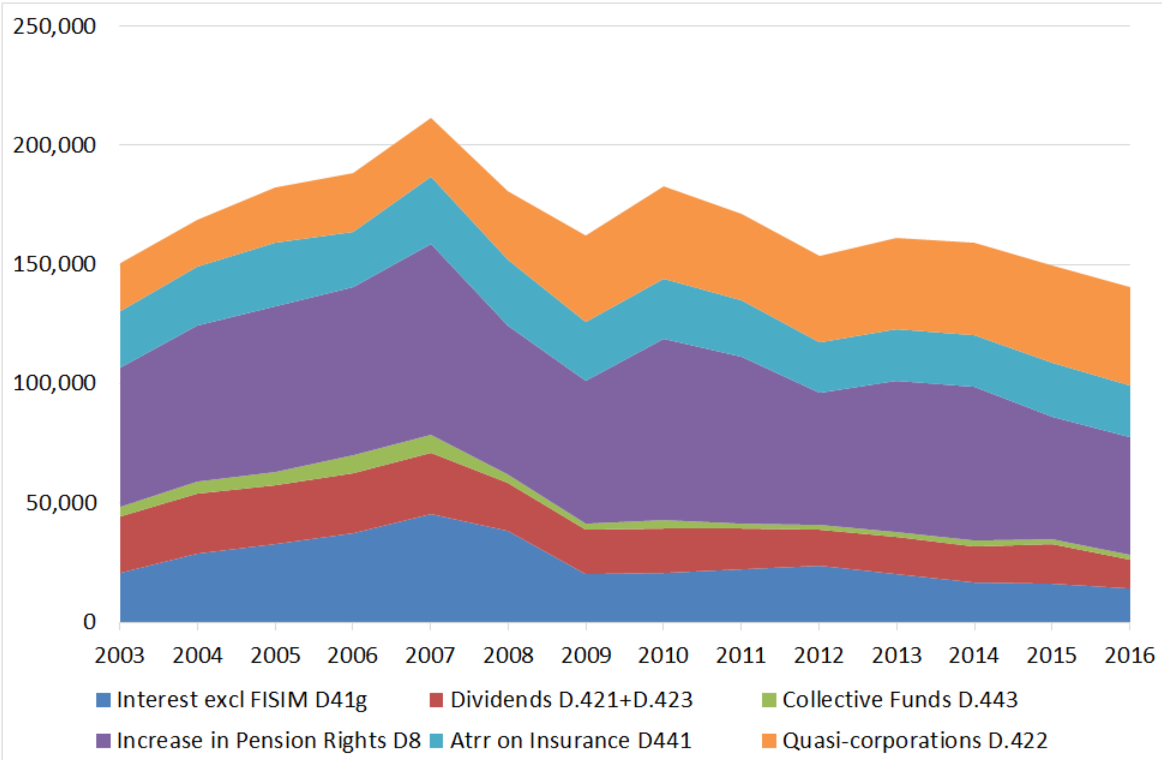
Component	**National Accounts Total (£mn)**	**LCFS Total (£mn)**	**Coverage Rate (%)**
Macro resources (receipts)			
Operating Surplus	130,150	68,060	52
Mixed Income	110,469	63,274	57
Wages and Salaries	711,054	663,206	93
Net Property Income Received	149,811	34,396	23
Social Benefits other than STiK	332,504	231,013	69
Social Transfers in Kind	273,509	179,603	66
A. Total	1,707,497	1,221,552	73
Macro uses (payments)			
Current Taxes on Income and Wealth	195,524	142,923	73
Employers act. Soc Contributions	136,091	59,606	44
Households' Social Contributions	67,528	62,945	93
B. Total	399,143	265,474	67
Household Disposable Income (A-B)	1,308,354	974,078	74
Memo: gross prop income excl rent	75,903	21,651	29

Source: Tonkin (2015) except totals and gross property income which is own calculations and excludes rent. The macro value is D.41g+D.42+D.443

In this table property income is shown net, while we are concerned with gross flows. Figure 2 shows the components of investment income identified in the national accounts. We aggregate interest, dividends and income from collective funds into a second category, investment income which we measure before any imputation for FISIM. We should expect these items to be recorded in the LCFS. The graph also shows the increase in pension rights. This is the variable that we seek to allocate across households; the treatment of pension funds, and the relationship of this term with investment income is explained in more detail below.

There are two "orphan" variables. Withdrawals from the income of quasi-corporations

Figure 2: Pension Accruals and Components of Household Investment Income (£mn)



Source: ONS data, July 2017.

are income accruing to partnerships, including limited liability partnerships. We assume that these are treated as self-employment income in the LCFS, and are not coded as investment income in the other surveys that we use. Income accruing on insurance products is not straightforwardly related to variables observed in any of our plausible survey sources and we therefore aggregate it with the increase in the value of pension rights as a single category to be allocated across households. The reasons for this are explained in more detail below.

3.2 The Treatment of Occupational Pensions in Distributional Accounts

The treatment of occupational pensions in distributional accounts needs to mirror that the System of National Accounts. There are two possible approaches each of which would be coherent. One would be to treat pension contributions and the investment income accruing to pension funds as a form of saving with household pension receipts as a corresponding form of dis-saving. If this were done it would not, however, also be possible to see pensions as a form of household current income. The alternative, adopted in the System of National Accounts, is therefore to treat pensions as social contributions with pension receipts seen as a form of social benefit in much the same light as state benefit income. This approach probably corresponds much more closely to the way in which people see their pensions, even if the reality is probably somewhere between the two, with regular receipts seen as income while lump-sum withdrawals are more like dis-saving. Since the 2015 Budget, of course, there are no rules limiting lump-sum withdrawals and it may, in many cases, not be possible to distinguish in any data regular from irregular withdrawals. But that is a general issue, and not one particularly associated with distributional accounts.

If private pensions are to be treated analogously to state-intermediated transfers, it then makes sense to treat any surpluses or deficits of pension funds in flow terms in the same way as the surplus or deficit of the government is treated. Piketty, Saez, and Zucman (2018) suggest that half of any government financial imbalance should be allocated to income and the other half to expenditure;[4] the same principle can be applied to pension funds. In the national accounts, however, an imputation is made to reflect any increase in pension liabilities of the corporate sector which is not matched by actual contributions and investment income. This item (D.612) has the consequence that the "surplus" of pension funds, instead of being the financial surplus is the increase in the value of pension rights (D8). Conversely when we come to assess the undistributed income of corporations, we measure it net of this imputation. In financial year 2013/14 the surplus of pension funds was £62.8bn and £13.1bn of this was imputed income.

The allocation is most sensibly made in proportion to the value of pension rights owned by each household. In practice this means, of course, that because each half of the adjustment is made in proportion to the same variable (in contrast to the government adjustment because the amount of any deficit allocated to taxes will be paid by different households from the amount allocated to expenditure), there is simply an allocation of the financial surplus of pension funds to households in proportion to their pension rights.

[4] They seem to treat, in contradiction to the System of National Accounts, all public procurement as consumption. Treatment of the public sector consistent with the SNA implies that public sector net saving rather than the financial surplus should be allocated on a 50/50 basis.

3.3 Pension Rights and Insurance Products

The Wealth and Assets Survey asks respondents detailed information about their pension rights, and provides an estimate of the total current value of these. Every attempt is made to ensure that the figures are comprehensive. It also provides an estimate of the total value of insurance products such as endowment policies. Just as the total value of pension products can be used to allocate investment income of pension funds, so too, the total value of insurance products can be used to impute income accruing to owners of insurance products. There are, however, some unresolved issues over the data. The total value of pension rights in WAS in 2012-2014 is shown as £4.45tn while in the national balance sheet household sector ownership of "insurance, pension and standardised guarantee schemes" is £3.71tn. WAS also shows £24bn of endowment products associated with mortgages, and £66bn of other insurance products. On the other hand the sectoral income accounts show an income attributed to insurance products of £21bn and one attributed to pension funds of £67bn. Plainly it is not possible for the £90bn of insurance products (including endowments) to generate an income of £21bn. Moreover the income flow allocated to insurance products, unlike the income flow allocated to pension rights, remains with the household sector, and is shown as a component of household disposable income.

As noted above, our focus is on allocating income to households rather than disposable income. Pending clarification of the split between insurance and pension products, we aggregate pensions and insurance, and impute total holdings of these to the households in the LCFS. We then allocate the total income flow associated with insurance products (item D.441) along with the surplus of pension funds (item D.8) in proportion to individual households' imputed holdings of pension and insurance assets.

3.4 Investment Income

The magnitude of the problem faced in developing satisfactory stochastic estimates of household investment income can be seen from table 2. This shows estimates of total investment income generated from four widely-used micro sources, the LCFS, the Wealth and Assets Survey (WAS, Office for National Statistics (2017)), the data set Households Below Average Income (HBAI, **?**) and finally the Survey of Personal Incomes

(SPI). The LCFS is a survey in which each responding household participates only once. Respondents are asked to provide details of their income from all sources; expenditure data are collected largely by means of an expenditure diary kept for two weeks. There are about 7,000 households participating in the survey. WAS is a panel survey designed to collect data on wealth but also asking information on income. The HBAI data set is produced by the Department of Work and Pensions using the results of the Family Resources Survey (FRS). Because it is felt that the FRS under-records people with high incomes, the records are augmented with data from the SPI. The SPI, is compiled as a random sample of tax records. Since this covers only people who pay income tax, it is not comprehensive but it provides a much better picture of high incomes than a survey source might be expected to. The tax threshold in 2013/4 for people aged under sixty-five was £9,440; the thresholds for older people were slightly higher. People who are only basic rate tax payers, those with incomes lower than £41,550 in 2012/3 are not required to complete tax returns. While they have tax records, tax on investment income is collected at source and is not linked to those records. For basic rate tax payers, therefore, the SPI is in turn enhanced by using survey data from HBAI so as to provide estimates. While the SPI should be the most comprehensive source, it omits property income received through tax-free Individual Savings Accounts since interest and dividends received by holdings in these does not have to be declared.

It is clear that the LCFS under-records investment income, while the coverage of WAS and HBAI is better. In order to impute combined dividend and interest income for households we use data from the Survey of Personal Incomes (SPI), splitting the sample into men and women. Given the origins of the SPI we look only at taxpayers.

Returning, however, to LCFS, table 3 shows that around 50% of taxpayers aged over 18 report zero income in the LCFS, compared to only around 20% in the SPI. This suggests that it would be undesirable to address the shortfall in LCFS (even relative to the Blue Book) simply by scaling. That will have the effect of preserving zeros and therefore, as we argued earlier, tend to overstate the increase in estimates of inequality associated with higher levels of investment income.

Table 2: Total investment income by data source (£000s)

	LCFS		WAS	HBAI		SPI		Bluebook
	all	taxpayers	all	all	taxpayers	all	taxpayers	all
2003	15,241,519	15,035,285		31,061,759	28,226,667	39,040,321	37,601,704	54,981,000
2004	15,857,633	15,585,669		34,669,450	31,987,409	49,107,541	47,568,225	73,444,000
2005	19,411,697	19,167,936		39,132,053	36,107,250	57,130,229	56,252,692	81,042,000
2006	20,801,278	20,524,743	21,952,261	42,940,689	39,428,476	63,439,144	62,539,342	90,551,000
2007	34,469,627	34,079,604		51,551,517	47,090,429	75,045,693	74,023,119	107,314,000
2008	19,728,484	19,219,977	17,957,485	47,157,256	42,230,341			101,314,000
2009	16,041,968	15,621,831		43,550,611	37,784,831	67,086,825	65,581,611	69,730,000
2010	13,896,471	13,252,108	23,509,925	42,166,420	37,888,128	49,825,993	47,778,451	55,750,000
2011	12,782,361	12,488,965		42,240,145	36,903,738			64,870,000
2012	17,795,473	17,510,585	30,833,312	42,283,598	37,068,608			67,185,000
2013	21,651,560	21,362,635		54,513,129	48,274,151	69,782,455	66,923,169	73,859,000
2014	15,800,924	15,347,896		53,882,302	48,645,836			84,824,000

Note: Fiscal year data except from WAS. Total investment income for all ages and all aged taxpayers only by data source. Identifying taxpayers in the WAS is not possible, and the taxpayer indicator in the 2001 SPI data for 2002, 2008, 2011, 2012 and 2014 are not currently available. *Source:* LCFS, WAS, and Bluebook from ONS; HBAI from DWP, and SPI from HMRC.

Table 3: Proportion receiving zero investment income in the LCFS and SPI

	LCFS			SPI		
	Report zero	Total	Proportion	Report zero	Total	Proportion
2001	11,118	22,099	0.50			
2002	10,454	21,604	0.48			
2003	13,112	32,599	0.40	6,609,181	28,488,670	0.23
2004	13,721	32,422	0.42	8,234,555	30,271,653	0.27
2005	13,652	31,928	0.43	8,756,789	31,060,064	0.28
2006	14,491	32,302	0.45	4,276,969	31,818,831	0.13
2007	15,139	33,234	0.46	4,886,722	32,458,692	0.15
2008	15,727	32,592	0.48			
2009	17,176	32,142	0.53	5,019,671	30,568,508	0.16
2010	18,266	32,068	0.57	8,392,544	31,289,990	0.27
2011	17,230	31,365	0.55			
2012	17,392	31,092	0.56			
2013	17,689	32,166	0.55	5,340,151	30,418,428	0.18
2014	12,478	23,649	0.53			

Note: Proportion of taxpayers aged 18 years and over who report receiving zero investment income in the LCFS and SPI. LCFS figures are in 000s. SPI data for 2001, 2002, 2008, 2011, 2012 and 2014 are not currently available. *Source:* LCFS from ONS; SPI from HMRC.

3.5 Taxation

The LCFS records payments of current taxes on income and wealth and, as table 1 makes clear, these are materially under-recorded. In contrast employees' social contributions (National Insurance contributions collected on employment income) are, like employment income itself, only modestly under-recorded relative to the national accounts. If, as we suggest, taxable income is to be imputed to individuals, to replace reported data for investment income, then it is clearly necessary for tax liabilities to be recalculated on the basis of imputed gross investment income. It would not seem sensible to impute income receipts on the basis of another data source and then to reconcile tax paid simply by means of scaling.

Programming in the basic system of allowances and tax rates in Stata is reasonably straightforward. Of course the final outcome has to depend on the way in which discrepancies other than investment income are closed and we do not, in this paper, address the production of adjusted estimates of household disposable income.

4 Imputation Methods and Models

In this section we present two means of modelling pension wealth. The first uses a simple selection model, making the assumption of normality. The second uses an ordered probit framework which makes much weaker distributional assumptions, but at the cost of requiring the data to be categorised. We show that the second approach is much more succcessful at reproducing the distribution of pension wealth observed in the Wealth and Assets Survey and we also use this method for modelling receipts of investment income. We then discuss the issue of covariances between the stochastic components of the data that we seek to impute.

4.1 A Simple Selection Model for Imputing Pension Wealth

In order to impute pension and insurance wealth, we break the WAS sample into four categories: i) households with heads aged under sixty-five who report employment income or self-employment income, ii) households with heads aged under sixty-five who do not report any employment or self-employment income, iii) households with heads aged sixty-five or more who report receiving private pension income and iv) households with heads aged sixty-five or more who do not report receiving any private pension income. All of the households in category iii) report owning pension assets while those in the other three categories may not own any pension/insurance assets. Households in category iv) may own pension rights but not be drawing any pension from them. Only wave 3 which covered the period 2010-2012 and wave 4, covering 2012-2014, provide adequate data on the components of income required for our model; we do not therefore attempt to estimate models for waves 1 and 2. The parameters estimated for wave 3 can certainly be used for years before 2010, but the greater the distance from the start of the survey, the greater must be the uncertainty surrounding the results.

Our first approach is to fit Heckman models using as explanatory variables age and age^2 reported in five-year bands (and numbered from one to seventeen), the number of adults and the number of children in each household. We also use dummies for housing tenure (owned outright, owned with a mortgage or other) and, in cases i) and iii) we use the log of employment plus self-employment income and the log of pension income as additional explanatory variables. The dependent variable, ln $y_{i,}$ is the log of insurance and pension wealth. In the absence of any obvious exclusion restriction, we use the same variables to estimate the selection equation, with the parameters being identified by the assumption that the disturbances are normally distributed. Thus the model is

$$\ln y_{ij} = X_{ij}\beta_j + \varepsilon_{ij}$$

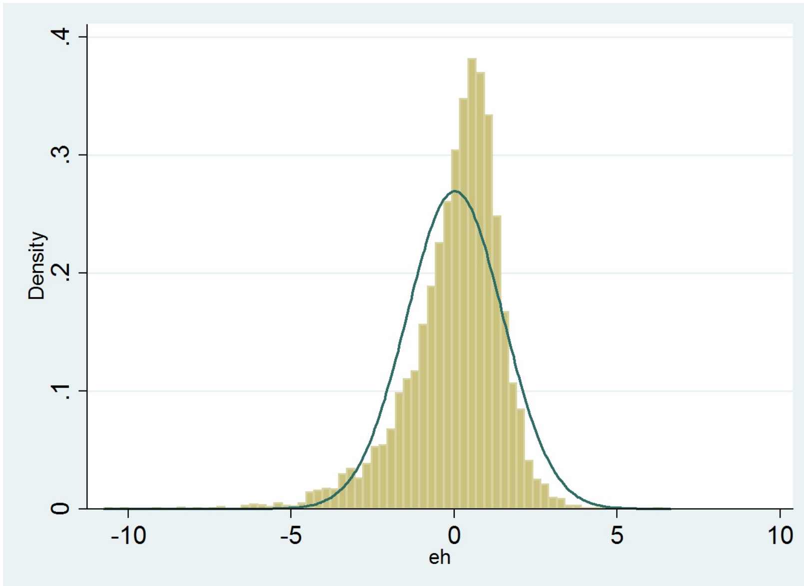
y_{ij} is observed if the latent variable

$$z_{ij} = X_{ij}\gamma_j + \eta_{ij} > 0, \quad \left[\begin{array}{c} \varepsilon_{ij} \\ \eta_{ij} \end{array} \right] \sim N\left(0, \left[\begin{array}{cc} \sigma^2 & \sigma\rho \\ \sigma\rho & 1 \end{array} \right]\right).$$

Here j ranges from 1 to 4 representing each of the four data categories and i indexes each observation within the category. The parameters of the model can be estimated using the heckman command in Stata. The parameters of this model are shown in table 5 and the selection equation is in table 6 in appendix A.

Figure 3 shows the histogram of the residuals from the Heckman equation fitted to wave 4 of WAS. It is clear that there is a very substantial departure from normality; a normal distribution has too many very high and low values, with not enough in the centre. This is a typical of a distribution with fat tails and, as we show subsequently, it makes the

Figure 3: The Residuals (actual minus predicted) of the Heckman Equation for Households with Heads under 65 and receiving Employment Income in Wave 4 of WAS



Heckman estimator unsuitable as a tool for imputation.

4.2 An Ordered Probit Model

The second approach is based on an ordered probit model. We allocate pension and insurance wealth to categories. We show in table 14 of appendix B the categories used, together with the number of observations in each, scaled up by the relevant weighting factors.

With N categories we allocate y_{ij} to category $K_{n,j}$ if it lies above the upper threshold for category $K_{n-1,j}$ denoted by $k_{n-1,j}$, and but is less than or equal to the upper threshold for category $K_{n,j}$ denoted by $k_{n,j}$, in other words if $k_{n-1,j} < y_{ij} \leq k_{n,j}$ where the category boundaries for k_1 to k_N are shown in table 14. k_0 can take any negative number since negative holdings of pension wealth cannot be observed and the treatment of the upper boundary ensures that with $k_1 = 0$ zero holdings are distinguished from strictly positive holdings. The ordered probit model assumes that, underlying each observation there is a latent variable

$$z_{ij} = X_{ij}\delta_{ij} + \nu_{ij}; \quad \nu_{ij} \sim N(0, 1)$$

and a set of thresholds $k^*_{1,j}$ to $k^*_{N-1,j}$ such that, if

$$k^*_{n-1,j} < z_{ij} \leq k^*_{n,j} \text{ then } y_{ij} \in K_{n,j}$$

This model can be estimated by means of maximum likelihood using the *oprobit* command in Stata. Parameter estimates for wave 4 are shown in table 7 and those for wave 3 in table 8 of appendix A.

4.3 The Ordered Probit Model and Extreme Values

As noted above, a common feature of both income and wealth distributions (see e.g. Jenkins (2017)) is that they show a large number of extreme values. In this section we explain how, if the underlying variable is distributed with a commonly-used extreme value distribution, the Pareto distribution of the first type, that relates to the ordered probit model we propose. The Pareto type-1 distribution for observations $x_i > x_m$ has the form

$$1 - F(x) = (x_m/x)^{\alpha} \text{ with } \alpha > 0$$

and the expected value conditional on $x > x_m$ is $x_m\alpha/(\alpha - 1)$ if $\alpha > 1$ but infinite otherwise. Fitting a Pareto distribution[5] to the data for the value of pension and insurance assets for those holdings greater than £2m, we find a value of $\alpha = 3.16$ for wave 3 and $\alpha = 2.66$ for wave 4. For imputed values greater than the highest cut point of £5mn we follow Jenkins (2017) and impute the expected value conditional on being greater than £5mn. This expected value is £7.31mn for values imputed using the distribution of wave 3 of WAS and £8.01mn for values imputed using wave 4 of WAS.

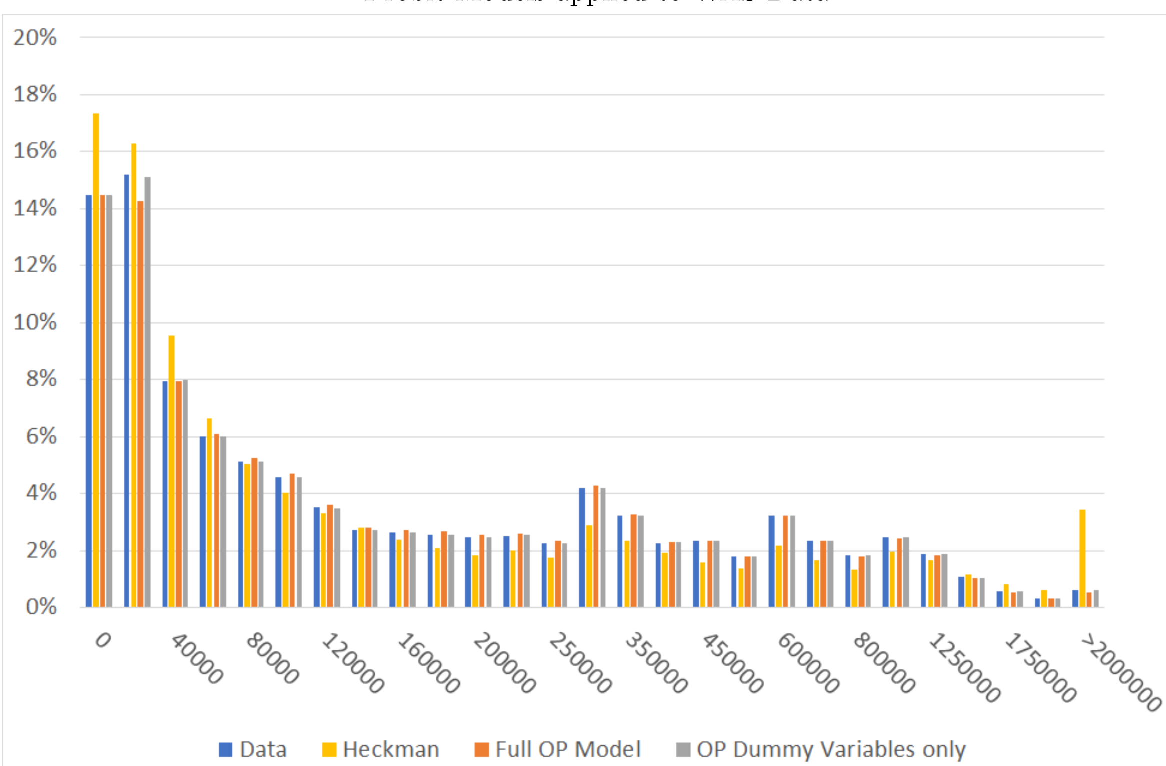
4.4 The Distribution of Simulated Data: the Heckman and Ordered Probit Models Compared

In figure 4 we show the distribution of the value of pension rights generated from the fourth wave of WAS, first using the Heckman and secondly with the ordered probit model; here they are fitted to unweighted data in order to maintain the transparency of the exercise. The results shown are generated as the average of five hundred distributions

[5] Using the Stata routine *paretofit* (see Jenkins and van Kearn (2015)).

of the relevant stochastic models. We find that the Heckman model overstates the proportions of the people with low pension rights, and also those right at the top of the distribution. It gives a picture of a distribution appreciably more uneven than it actually is. The full ordered probit model under-records the proportion with non-zero, but small pension pots but is otherwise close to the data. Our final model shows the ordered probit model estimated only with dummy variables for marital status and housing tenure. Here the fit of the distribution to the actual data is very good (as it is if the distribution is modelled without the use of any covariates). This suggests that the deviation of the expected distributional shape from the data arises from the use of continuous variables (age, age^2 and log employment/self employment income) as covariates. A similar issue appears much more markedly with the proportion of zeros implied by the Heckman estimator. The high simulated proportion of low and very high pension rights is probably not a consequence of the need to identify the model parameters by means of the assumption of normality; rather figure 3 suggests it is the outcome of the assumption of normality in the simulation exercise. We make use of the ordered probit models of tables 7 and 8 in our subsequent work

Figure 4: The Distribution of Pension Rights simulated for 2013 using Heckman and Ordered Probit Models applied to WAS Data



The Wealth and Assets Survey does not provide comprehensive income data in waves 1 and 2; we therefore fit the ordered probit model to waves 3 and 4, using wave 4 parameters for 2012-2014 and wave 3 parameters for earlier years.

4.5 An ordered probit model of receipts of investment income

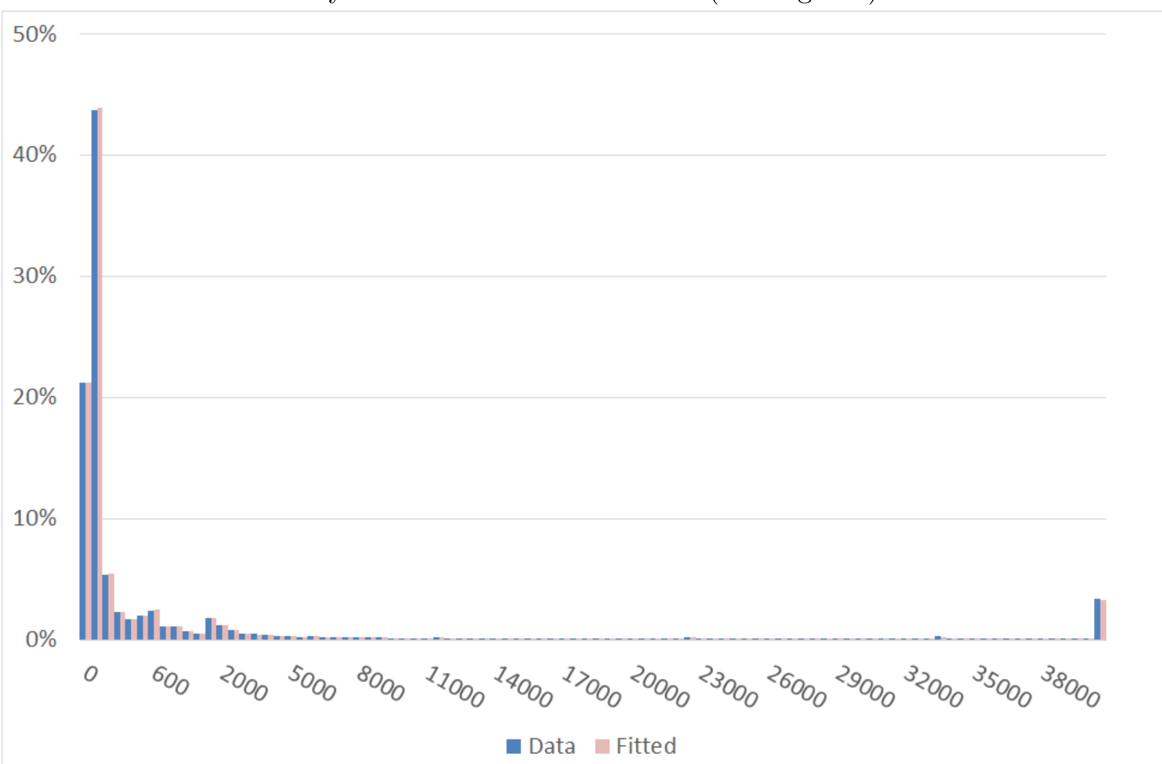
We similarly use an ordered probit model to allocate investment income to categories. Table 15 in appendix B shows the categories used together with the observations in each in 2013, scaled up by the relevant weighting factors. The table shows that 18% of taxpayers in the SPI in 2013 reported investment income of zero, while 53% reported income of between £0 and £100. Approximately 1% of respondents reported investment income of more than £40,500, although in value terms this accounted for 47% of the total investment income reported. The model is outlined above, and the explanatory variables we use are age bands (6), the log of labour income, and regional dummies.[6] We estimate the model separately for men and women, and separately for each year before translating the coefficients into the LCFS for imputation. The estimated coefficients from the SPI are shown in table 9 for men and table 10 for women in Appendix. A[7] It is notable that the coefficients on log of income excluding investment income[8] (*linc*) are universally negative for women and at best only weakly positive for men. This suggests that investment income is a substitute for rather than a complement to labour income. One explanation may be that a high proportion of people who are in practice self-employed have chosen to incorporate and receive their income as dividends rather than as self-employment income appearing in *linc.* Older people are more likely to receive investment income, and receipts are also likely to be higher in regions such as London (region 7) and the South-East (8) relative to the North-East (omitted).

A small number of people have no reported income excluding their investment income. For these, again distinguishing men and women, we estimate models with only age and regional dummies present. These show similar age and regional patterns similar to the bulk of the population with reported non-investment income and the results of this are also presented in appendix A.

[6]The SPI is limited in having few characteristics of individuals in it.
[7]The cut points are not shown, but are available on request.
[8]This includes pensions and self-employment income.

Given that between 30-50% of total investment income receipts accrue to people receiving more than £40,500 p.a., we use the Pareto distribution as discussed above to model the large number of extreme values of investment income. We estimate the Pareto parameter α from the SPI as a function of time and sex, with the parameters shown in table 13 in the appendix. Figure 5 shows the distribution of the value of actual investment income for 2013 and the predictions using the ordered probit model in the SPI. The figure shows that the fit of the distribution to the actual data is very good. A similarly good fit is obtained with the models estimated on weighted data.

Figure 5: The Distribution of Investment Income in the 2013 SPI and the Distribution Fitted by the Ordered Probit Models (Unweighted)



We explored fitting a Heckman model as an alternative, but the nature of the data was such that we did not find coherent estimates. This is probably a consequence of having to rely on the assumption of normality in order to identify the coefficients.

4.6 Covariances

We have so far set out the stochastic models which we use to impute estimates of pension wealth, dividend and interest receipts and interest payments to the Living Costs and

Food Survey. These models are estimated using three different and unrelated sources, and the essence of stochastic imputation is that there is a random element to each imputed variable. This means that, while we cannot give any meaning to any particular record, we can hope to improve our understanding of the distribution of income. But the implications for household income of these imputations will depend very much on the extent to which the random components of the imputed variables are correlated. If we were to assume them not to be correlated while, in fact there are positive correlations on the receipts side, we would probably tend to understate the underlying inequality in incomes. Since, however, we have tried to use the most suitable source for each imputed item, we are unable to estimate our models jointly so as to provide estimates of the correlations simultaneously with the model parameters. Furthermore, in our categorical models we have divided the categories finely. This has merits in that it improves our representation of the underlying distributions. But, at the same time, the high number of parameters makes any joint estimation difficult.

The only source which allows any form of joint estimation is the Wealth and Assets Survey. As well as the data on pension rights for which it is the only source, it provides information on other financial assets. These are provided on an individual basis allowing them to be related to the individual data in the SPI. We model jointly, the distribution of pension rights by households with the value of financial assets owned by the household reference person and her/his spouse/partner. Making the assumption that receipts of dividends and interest are proportional to financial assets, this then allows us to estimate the correlation matrix for the random components to the latent variables behind household pension rights and dividend and interest receipts of the household reference person and spouse.

The model we estimate specifies eight categories for pension assets and nine for other financial assets. The equations we use correspond to those described above, with the exception that, since we can estimate the model only for households where there is a household reference person and spouse/partner, we do not include dummies representing marital status.

From a given wave of the WAS we assume, with i indexing the observations, j indexing the subcategory of WAS data (employment income and age of household reference person<65, no employment income and age of household reference person<65, pension income and age of household reference person 65+, no private pension income and age of household reference person 65+) and k indexing each of the three variables of interest (assets of the reference person, assets of their spouse, and the pension rights of the household), that there is a latent variable defined as

$$z_{ijk} = X_{ijk}\delta_{ijk} + \nu_{ijk}$$

with thresholds $k^*_{1,j,k}$ to $k^*_{N-1,j,k}$ and categories $K_{n,j,k}$ such that, if $k^*_{n-1,j} < z_{ij} \leq k^*_{n,j}$ then $y_{ijk} \in K_{n,j,k}$

Here

$$\begin{bmatrix} \nu_{ij1} \\ \nu_{ij2} \\ \nu_{ij3} \end{bmatrix} \sim N \left(0, \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix} \right)$$

where the correlations shown by the $\rho_{m,n}$ are the focus of our interest. ρ_{12} is the correlation between the latent variables driving investment income for each of the two adults; ρ_{13} is the correlation between the latent variables driving investment income of the first adult and that driving pension rights; and ρ_{23} is the correlation between the latent variables driving investment income of the second adult and that driving pension rights. For each of the four categories, then, it is necessary to estimate a multivariate ordered probit model, with, of course, the risk that it may not be possible to estimate the model in all cases.

Except for the correlation between assets held by the household reference person and their partner, the correlations are generally small. In order to simplify the programming, we use a single set of correlations for all of the four WAS subcategories. Details of the correlations are provided in table 4 for those correlations which could be estimated.

Table 4: Estimates of the Correlation Matrix

	Wave 3						Wave 4				
	<65 Empl Inc.		<65 No Empl Inc		>64 Pens. Inc.		<65 Empl. Inc.		<65 No Empl. Inc		Mean
	ρ	S.E.	ρ	S.E.	ρ	S.E.	ρ	S.E.	ρ	S.E.	
ρ_{12}	0.78	0.01	0.88	0.01	0.80	0.01	0.78	0.01	0.88	0.01	0.82
ρ_{13}	0.24	0.01	0.42	0.04	0.10	0.02	0.23	0.01	0.43	0.04	0.28
ρ_{23}	0.25	0.01	0.47	0.04	0.08	0.02	0.22	0.01	0.44	0.04	0.29

5 Simulation

We use our ordered probit models to impute values for pension rights and interest and dividend receipts. A fitted value for each latent variable is computed, and random terms drawn from the multivariate normal distribution with correlations as indicated by table 4 are added on. Each latent variable is then allocated to the relevant category underpinning the ordered probit model. Where it lies between two cut points, the distance between the two categorical boundaries is interpolated on the basis of the latent variable relative to the distance between the two cut points. Zeros, however, remain zeros and, as we have noted above, very high values are imputed using the Pareto distribution. We do, however, impute values for pension rights of less than £20,000 using a uniform distribution in the range [0 20,000], because this corresponds better to the observed distribution than does the outcome delivered by our more general imputation approach.

For pension rights it is now straightforward to allocate the sum of the surplus accruing to pension funds and the investment income associated with insurance products in proportion to pension rights (European system of Accounts codes D.8 and D.441). Our imputation of interest and dividends is applied only to tax-paying individuals. *Faut de mieux* we use the existing LCFS figures for non-taxpayers.

Dividends and interest are subject to taxation, while the taxation of income on insurance products is unclear, and the increase in the value of pension rights is not taxed, although of course taxes fall due when pensions are paid. We calculate the change in gross income which results from our imputations of interest and dividends. As explained earlier, given the excess of SPI figures over the aggregate data, we have not at this stage proceeded to reconcile our figures with the Blue Book, or to allocate undistributed corporate income. Using variable P051 in the LCFS as gross income before the imputations, we add on any imputed investment income for taxpayers, but deduct any shown in the survey. As described in section 3.5 we calculate the individual's tax bill before and after the change in the imputation, and, in the individual records, also store investment income net of tax.

This then allows us to examine the effect on standard measures of inequality of our imputations for under-recording of investment income and from the exclusion of income associated with pension rights and insurance policies. The results, however, inevitably vary from one realisation to another. We undertake repeated simulations, showing the mean and standard deviation of the Gini coefficient associated with i) income with investment income aligned with the Blue Book, and ii) income also taking account of returns to pension rights and insurance products. We also show the impact of our changes on the geometric mean of equivalised household income since developing good estimates of this variable is the underlying motive of our work. We are in essence drawing on the proof presented by Rubin (1987) that, provided the simulations are generated using valid probabilistic models, the normal principles of statistical inference can be applied to data such as the Gini coefficients and estimates of the geometric mean derived from them. Here we present results drawn from five simulations; these are adequate to give a sense of the dispersion of the aggregates derived from simulated data.

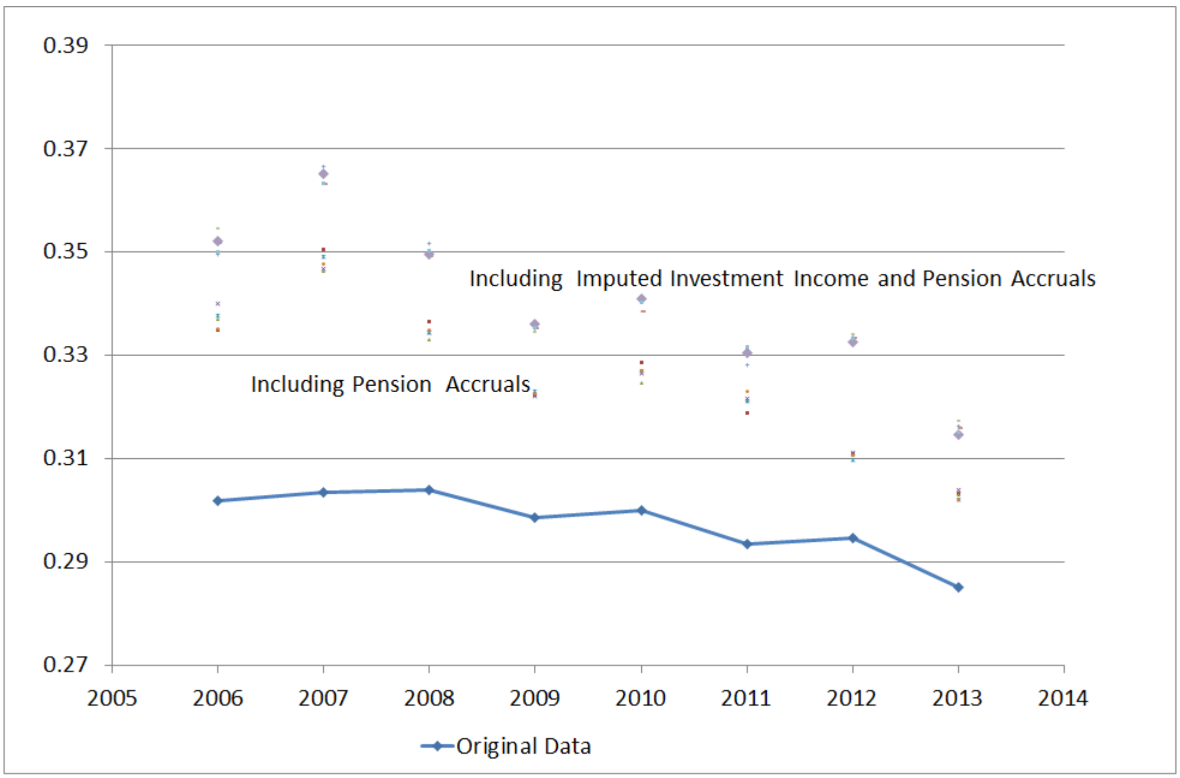
The results presented here are produced using the publicly-available LCFS because development work using this is much more convenient than using the full version with controlled access. In the public version incomes are top-coded[9] so as to reduce the risk of identifying respondents. That means, particularly with pension rights where our models for households of working age are strongly sensitive to labour income, that we do not reproduce the top end of the distribution. Pension incomes in the LCFS are, in contrast not top coded, and the same problem does not arise for those households with heads over sixty-five who receive pension income. Similarly, the influence of labour income on investment income was weak (and sometimes negative). As a result imputation of investment income should not be greatly affected by top coding.

Figure 6 shows the Gini coefficient for equivalised household income calculated from the top-coded gross gross income data in LCFS, but with taxes deducted as explained above. It also shows two groups of stochastic simulations, the first for income including pension fund accruals, and the second with both pension fund accruals and investment income for taxpayers imputed, and tax payments adjusted for the change in tax liabilities resulting from the latter. While it will be desirable to carry out further simulations, the chart gives the impression that stochastic simulation will probably not lead to a high level of uncertainty. The actual levels for the Gini coefficient, and the subsequent values we show for the geometric mean of household income are, of course, very materially affected by top coding, and no attention should therefore be paid to these.

[9]Although the investment income and pension income figure which go to produce these are not top-coded.

Figure 7 shows comparable figures for the geometric mean of household income. Again, in addition to the original data, there are two groups of five stochastic simulations each. The lower group shows the geometric mean of income including pension fund accruals, while the higher group shows the geometric mean of income including, in addition, imputed investment income for taxpayers. It can be seen that, while the additional inclusion of imputed investment income raises the Gini coefficient materially, it does very little to increase the geometric mean of household income. The reasons for this can be inferred from a comparison of figures 2 and 5. It is clear that most recipients of investment income receive only very small amounts, while pension rights are much more evenly spread among the population.

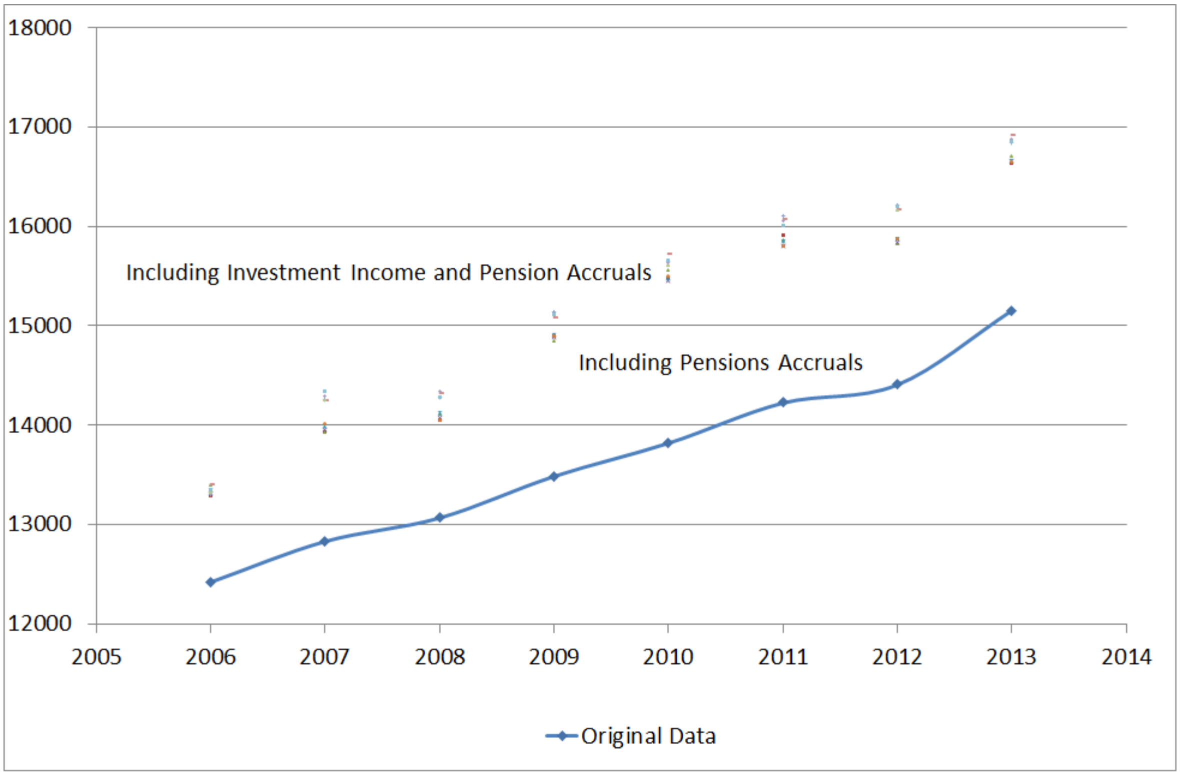
Figure 6: Estimates of the Gini Coefficient with Different Definitions of Income: 2006-2013



6 Conclusions and Future Work

We have demonstrated here the practicality of using stochastic methods first to adjust income data in the Living Costs and Food Survey for accrual of pension rights, and secondly to impose the pattern of investment income receipts for taxpayers shown in the

Figure 7: The Geometric Mean of Equivalised Household Income (£ p.a.) with Different Definitions of Income



Survey of Personal Incomes. In the data we have each of these is shown to be a source of increased inequality, but it remains to be established how for this is a consequence of the fact that we have, so far, been working with top-coded data. Having developed models and the programmes needed to apply them to the Living Costs and Food Survey, we are now in a position to implement them in secure circumstances with data sets which are not top coded.

The model presented here has combined interest and dividend income. Having developed a working model for this, we need to split the two sources of income. Preliminary work suggests that dividend receipts are more strongly negatively correlated with labour/pension income than is the total, perhaps accounting, at least to some extent, for the very low reported incomes of some households in the LCFS.

7 References

Brewer, Mike, Ben Etheridge, and Cormac O'Dea. 2017. "Why are Households that Report the Lowest Incomes So Well-off?" *The Economic Journal* 127 (605):F1–F631.

Crane, Eric. 2017. "Home Economy National accounts UK sector accounts National Accounts articles Article: National Accounts articles: Improvements to the treatment of Corporate Bonds Interest." Tech. rep. URL https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/articles/nationalaccountsarticles/improvementstothetreatmentofcorporatebondsinterest.

Crossley, Thomas F. and K Pendakur. 2010. "The Common-Scaling Social Cost-of-Living Index." *Journal of Business and Economic Statistics* 28 (4):523–538.

Digital Economy Act. 2017. "Digital Economy Act. 2017." URL Www.legislation.gov.uk/ukpga/2017/30/section/8/enacted.

Fesseau, M. and M.L. Matteonetti. 2013. "Distributional Measures across Household Groups in a National Accounts Framework." URL http://dx.doi.org/10.1787/5k3wdjqr775f-en.

Fixler, D. and D.S. Johnson. 2014. "Accounting for the Distribution of Income in the U.S. National Accounts." In *Measuring Economic Sustainability and Progress*, edited by D.W. Jorgenson, J.S. Landefeld, and P. Schreyer. Chicago: Chicago University Press, 213–244.

Fleurbaey, M. 2009. "Beyond GDP: the Quest for Measures of Social Welfare." *Journal of Economic Literature* 47 (4):1029–75.

HM Revenue and Customs. 2016. "Survey of Personal Incomes 2000-01 to 2013-14. Public Use Tape." URL http://doi.org/10.5255/UKDA-SN-{%}0A8044-1.

Jenkins, S P. 2017. "Pareto Models, Top Incomes and Recent Trends in UK Income Inequality." *Economica* 84 (334):261–289.

Jenkins, S P and P. van Kearn. 2015. "Pareto t: Stata Module to t a Type 1 Pareto Distribution." URL https://ideas.repec.org/c/boc/bocode/s456832.html.

Jorgenson, D. and P. Schreyer. 2015. "Measuring Individual Economic Well-Being and Social Welfare within the Framework of the System of National Accounts."

Kent, K. 2017. "National Accounts articles: an update to assessment of changes to sector and nancial accounts, 1997 to 2015." Tech. rep. URL https://www.ons.gov.uk/releases/nationalaccountsarticlesanupdatetoassessmentofchangestosectorandfinancialaccounts1997t

Layard, R., S. Nickell, and G. Mayraz. 2008. "The Marginal Utility of Income." *Journal of Public Economics* 92 (8-9):1846–1857.

Lewbel, Arthur. 1997. "Consumer Demand and Household Equivalence." In *Handbook of Applied Economics*, edited by H. Pesaran and P. Schmidt. Blackwell Publishers.

Meade, J E and J R N Stone. 1941. "The Construction of Tables of National Income, Expenditure, Savings and Investment." *Economic Journal* 56 (202/203):216–233.

Office for National Statistics. 2017. "Wealth and Assets Survey, Waves 1-4." URL http://doi.org/10.5255/UKDA-SN-7215-5.

Office for National Statistics and Dept for Environment, Food and Rural Affairs. 2017. "Living Costs and Food Survey." URL http://doi.org/10.5255/UKDA-SN-8210-1.

Piketty, Thomas, Emmanuel Saez, and G Zucman. 2018. "Distributional National Accounts: Methods and Estimates for the United States." *Quarterly Journal of Economics* :1–57.

Rubin, D B. 1987. *Multiple Imputation for Non-response in Surveys*. New York: Wiley.

Saez, Emmanuel and G Zucman. 2016. "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data." *Quarterly Journal of Economics* 131 (2):519–578.

Schafer, J L. 1999. "Multiple Imputation: a Primer." *Statistical Methods in Medical Research* 8 (1):3–15.

Sefton, J and M R Weale. 2006. "The Concept of Income in a General Equilibrium." *Review of Economic Studies* 73 (1):219–249.

Stiglitz, Joseph E., Amartya Sen, and J-P Fitoussi. 2009. "Report by the Commission on the Measurement of Economic Performance and Social Progress." Tech. rep., INSEE, Paris.

Stone, R and G Stone. 1972. *National Income and Expenditure*. London: Bowes and Bowes.

Tonkin, Richard. 2015. "The Distribution of Household Income, Consumption and Savings: an OECD Study." Tech. rep. URL https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/articles/thedistributionofhouseholdincomeconsumptionandsavingsanoecdstudy/2015-11-30.

van Buuren, S. 2007. "Multiple Imputation of Discrete and Continuous Data by Fully Conditional Speci cation." *Statistical Methods in Medical Research* 16 (3):219–242.

A Models used to Impute Missing Data

Table 5: The Parameters of the Linear Component of the Heckman Model: Wave 4 of the Wealth and Assets Survey

	Under 65		65 and Older	
main	Emp Inc	No Emp Inc	Pens Inc	No Pens Inc
age	1.045***	0.072	2.720***	6.798
	(0.109)	(0.330)	(0.259)	(4.726)
age^2	-3.278***	1.639	-10.032***	-21.752
	(0.563)	(1.532)	(0.854)	(15.333)
No. Adults	-0.163***	-0.055	0.184***	-0.823*
	(0.029)	(0.090)	(0.026)	(0.373)
No. Children	-0.008	0.04	-0.103	0.165
	(0.022)	(0.081)	(0.055)	(0.381)
Hse owned outright	1.005***	1.379***	0.047	-0.653
	(0.069)	(0.121)	(0.030)	(0.470)
Hse mortgaged	0.675***	0.847***	0.077	-1.181
	(0.059)	(0.129)	(0.051)	(0.727)
Married etc	0.391***	0.711***	-0.147***	0.662
	(0.074)	(0.145)	(0.038)	(0.428)
Single	0.177*	0.052	-0.083*	-0.37
	(0.087)	(0.147)	(0.041)	(0.723)
Widowed	0.523**	-0.075	-0.059	-0.687
	(0.160)	(0.221)	(0.031)	(0.464)
ln Lab Income	0.746***			
	(0.031)			
ln Pens Income			0.974***	
			(0.010)	
Constant	-4.159***	7.161***	-15.115***	-37.663
	(0.557)	(1.749)	(1.969)	(36.175)

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: The Selection Equations of the Heckman Model:
Wave 4 of the Wealth and Assets Survey

	Under 65		65 and Older	
main	Emp Inc	No Emp Inc	Pens Inc	No Pens Inc
age	0.411***	0.474***		-2.949
	(0.084)	(0.134)		(1.514)
age^2	-1.514***	-1.742*		8.778
	(0.459)	(0.681)		(4.905)
No. Adults	-0.137***	0.047		0.785***
	(0.026)	(0.063)		(0.099)
No. Children	-0.051*	-0.154***		-0.540**
	(0.023)	(0.038)		(0.182)
Hse owned outright	0.898***	1.234***		0.510***
	(0.075)	(0.091)		(0.106)
Hse mortgaged	0.617***	1.002***		1.094***
	(0.047)	(0.090)		(0.208)
Married etc	0.296***	0.195		-0.248
	(0.064)	(0.103)		(0.172)
Single	0.13	-0.156		0.076
	(0.074)	(0.086)		(0.227)
Widowed	0.144	0.091		0.279
	(0.149)	(0.144)		(0.158)
ln Lab Income	0.523***			
	(0.030)			
Constant	-7.123***	-3.368***		21.937
	(0.437)	(0.642)		(11.646)
N	9968	2841		1574
N Censored	1442	1230		1273
	0.05**	0.05		0.86**
rmse	1.48	1.6		2.61

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: The Parameters of the Ordered Probit Model:
Wave 4 of the Wealth and Assets Survey

	Under 65		65 and Older	
main	Emp Inc	No Emp Inc	Pens Inc	No Pens Inc
	Under 65		65 and Older	
	Emp Inc	No Emp Inc	Pens Inc	No Pens Inc
age	0.401***	0.323**	2.888***	-2.735
	(0.056)	(0.121)	(0.499)	(1.421)
age^2	-0.682*	-0.717	-11.924***	7.993
	(0.303)	(0.600)	(1.638)	(4.620)
No. Adults	-0.119***	0.003	0.385***	0.554***
	(0.018)	(0.051)	(0.057)	(0.080)
No. Children	-0.014	-0.136***	-0.292*	-0.380**
	(0.013)	(0.034)	(0.129)	(0.137)
Hse owned outright	0.884***	1.331***	0.109*	0.442***
	(0.044)	(0.071)	(0.051)	(0.102)
Hse mortgaged	0.550***	0.889***	0.213*	0.961***
	(0.032)	(0.066)	(0.096)	(0.161)
Married etc	0.346***	0.398***	-0.200*	0.019
	(0.044)	(0.080)	(0.085)	(0.159)
Single	0.160**	-0.061	-0.117	0.035
	(0.051)	(0.070)	(0.075)	(0.202)
Widowed	0.270*	0.048	-0.032	0.271
	(0.116)	(0.109)	(0.063)	(0.152)
ln Lab Income	0.600***			
	(0.023)			
ln Pens Income			2.144***	
			(0.082)	
N	9968	2841	5857	1574
Psuedo-R^2	0.11	0.14	0.36	0.13

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Details of cut points are available on request

Table 8: The Parameters of the Ordered Probit Model: Wave 3 of the Wealth and Assets Survey

	Under 65		65 and Older	
main	Emp Inc	No Emp Inc	Pens Inc	No Pens Inc
age	0.577***	0.231*	3.749***	-1.582
	(0.055)	(0.113)	(0.546)	(1.402)
age^2	-1.501***	-0.287	-14.781***	4.351
	(0.295)	(0.582)	(1.822)	(4.517)
No. Adults	-0.141***	-0.215***	0.651***	0.492***
	(0.018)	(0.063)	(0.074)	(0.090)
No. Children	-0.031*	-0.100**	-0.320*	-0.325**
	(0.013)	(0.038)	(0.126)	(0.117)
Hse owned outright	0.850***	1.365***	0.140*	0.458***
	(0.044)	(0.078)	(0.059)	(0.096)
Hse mortgaged	0.597***	1.169***	0.149	0.925***
	(0.033)	(0.084)	(0.084)	(0.190)
Married etc	0.477***	0.622***	-0.549***	0.011
	(0.041)	(0.094)	(0.098)	(0.153)
Single	0.198***	0.069	-0.02	0.108
	(0.046)	(0.072)	(0.086)	(0.255)
Widowed	0.054	0.022	0.06	0.387**
	(0.094)	(0.110)	(0.070)	(0.127)
ln Lab Income	0.513***			
	(0.021)			
ln Pens Income			2.212***	
			(0.099)	
N	11525	2483	5586	1549
Psuedo-R^2	0.11	0.15	0.37	0.11

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Details of the cut points are available on request

Table 9: The Ordered Probit Models used to fit Investment Income in the SPI:
Men, Labour Income data available

	2003	2004	2005	2006	2007	2009	2010	2013
oinvtota								
Age 25-34	-0.013	0.137***	0.133***	0.071***	0.075***	-0.001	0.016	0.229***
	(0.010)	(0.014)	(0.013)	(0.011)	(0.012)	(0.011)	(0.012)	(0.014)
Age 35-44	0.136***	0.847***	0.877***	0.440***	0.623***	0.562***	0.710***	0.688***
	(0.011)	(0.013)	(0.012)	(0.010)	(0.011)	(0.011)	(0.011)	(0.014)
Age 45-54	0.275***	0.913***	0.952***	0.520***	0.722***	0.648***	0.805***	0.781***
	(0.011)	(0.013)	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)	(0.014)
Age 55-64	0.446***	1.007***	1.045***	0.649***	0.843***	0.739***	0.934***	0.897***
	(0.011)	(0.013)	(0.012)	(0.010)	(0.011)	(0.011)	(0.011)	(0.014)
Age 65+	0.912***	1.162***	1.110***	0.665***	0.961***	0.877***	1.035***	1.059***
	(0.011)	(0.014)	(0.013)	(0.011)	(0.011)	(0.010)	(0.011)	(0.014)
Ln Earned.	0.035***	0.115***	0.110***	0.091***	0.057***	-0.023***	-0.074***	-0.043***
Income	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
North-West	0.024	0.021	0.054***	0.019	0.035**	-0.000	0.029*	0.048***
	(0.014)	(0.014)	(0.014)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Yorks/Humbs	0.033*	0.020	0.041**	0.008	0.021	-0.017	0.005	0.034**
	(0.015)	(0.015)	(0.014)	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)
E. Midlands	0.057***	0.036*	0.071***	0.040**	0.061***	0.019	0.039**	0.050***
	(0.015)	(0.015)	(0.014)	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)
W. Midlands	0.031*	0.021	0.065***	0.024	0.048***	0.001	0.029*	0.053***
	(0.015)	(0.015)	(0.014)	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)
E. England	0.068***	0.050***	0.083***	0.036**	0.064***	0.018	0.053***	0.061***
	(0.014)	(0.014)	(0.014)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
London	0.081***	0.064***	0.097***	0.034**	0.069***	0.030*	0.078***	0.072***
	(0.014)	(0.014)	(0.014)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
South-East	0.097***	0.075***	0.109***	0.070***	0.111***	0.060***	0.097***	0.102***
	(0.014)	(0.014)	(0.013)	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)
South-West	0.093***	0.067***	0.083***	0.064***	0.087***	0.040***	0.073***	0.069***
	(0.015)	(0.014)	(0.014)	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)
Wales	0.016	-0.023	0.027	-0.017	0.030*	-0.041**	-0.013	-0.015
	(0.017)	(0.016)	(0.016)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Scotland	0.016	0.019	0.053***	0.027*	0.031*	-0.002	0.017	0.050***
	(0.015)	(0.015)	(0.014)	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)
N.Ireland	-0.082***	-0.071***	-0.062**	-0.090***	-0.051**	-0.137***	-0.065***	-0.092***
	(0.021)	(0.021)	(0.019)	(0.018)	(0.018)	(0.017)	(0.017)	(0.017)
N	230201	289787	292926	312710	324865	301874	348196	318559
pseudo R-sq	0.020	0.032	0.031	0.013	0.023	0.022	0.032	0.021

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Details of the cut points are available on request

Table 10: The Ordered Probit Models used to fit Investment Income in the SPI:
Women, Earned Income data available

	2003	2004	2005	2006	2007	2009	2010	2013
oinvtota								
Age 25-34	0.068***	0.172***	0.128***	0.030*	-0.026	-0.055***	-0.028*	0.194***
	(0.013)	(0.016)	(0.015)	(0.012)	(0.013)	(0.014)	(0.014)	(0.014)
Age 35-44	0.274***	1.095***	1.082***	0.358***	0.674***	0.695***	0.655***	0.685***
	(0.013)	(0.014)	(0.014)	(0.011)	(0.012)	(0.013)	(0.014)	(0.014)
Age 45-54	0.337***	1.128***	1.141***	0.415***	0.728***	0.766***	0.733***	0.788***
	(0.013)	(0.015)	(0.014)	(0.011)	(0.012)	(0.013)	(0.013)	(0.013)
Age 55-64	0.455***	1.198***	1.203***	0.497***	0.828***	0.857***	0.877***	0.920***
	(0.014)	(0.015)	(0.014)	(0.011)	(0.012)	(0.013)	(0.014)	(0.014)
Age 65+	1.144***	1.230***	1.180***	0.360***	1.162***	1.153***	1.244***	1.290***
	(0.015)	(0.018)	(0.016)	(0.014)	(0.013)	(0.013)	(0.014)	(0.014)
Ln Earned	-0.116***	-0.094***	-0.058***	-0.093***	-0.104***	-0.209***	-0.315***	-0.124***
Income	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
North-West	0.050*	0.029	0.028	0.037*	0.060***	0.064***	0.041**	0.065***
	(0.020)	(0.017)	(0.017)	(0.016)	(0.016)	(0.015)	(0.016)	(0.015)
Yorks/Humbs	0.040	0.035	0.022	0.043**	0.049**	0.042**	0.043**	0.053***
	(0.021)	(0.018)	(0.018)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
E. Midlands	0.054**	0.026	0.070***	0.055***	0.064***	0.038*	0.027	0.069***
	(0.021)	(0.018)	(0.018)	(0.017)	(0.017)	(0.016)	(0.017)	(0.016)
W. Midlands	0.063**	0.036*	0.050**	0.072***	0.085***	0.075***	0.051**	0.082***
	(0.020)	(0.018)	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
E. England	0.093***	0.093***	0.093***	0.113***	0.119***	0.096***	0.094***	0.129***
	(0.020)	(0.018)	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.015)
London	0.107***	0.110***	0.134***	0.138***	0.132***	0.122***	0.133***	0.173***
	(0.019)	(0.017)	(0.017)	(0.016)	(0.016)	(0.015)	(0.015)	(0.015)
South-East	0.108***	0.106***	0.108***	0.126***	0.147***	0.125***	0.125***	0.154***
	(0.019)	(0.017)	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
South-West	0.097***	0.074***	0.077***	0.112***	0.117***	0.075***	0.089***	0.101***
	(0.021)	(0.018)	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Wales	0.049*	0.021	0.022	0.022	0.043*	0.034	-0.008	0.020
	(0.023)	(0.020)	(0.020)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Scotland	0.036	0.017	0.016	0.056***	0.053***	0.058***	0.030	0.069***
	(0.020)	(0.018)	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.015)
N.Ireland	-0.008	0.021	-0.025	-0.003	0.013	-0.006	-0.013	-0.026
	(0.026)	(0.024)	(0.023)	(0.022)	(0.022)	(0.021)	(0.022)	(0.020)
N	118015	148615	152765	163330	172163	186320	186933	232910
pseudo R-sq	0.031	0.040	0.040	0.008	0.032	0.040	0.051	0.035

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Details of the cut points are available on request

Table 11: The Ordered Probit Models used to Fit Investment Income in the SPI: Men, Labour Income not available

	2003	2004	2005	2006	2007	2009	2010	2013
oinvtota								
Age 25-34	-0.275	-0.235	-0.178	0.289	0.137	-0.153	-0.240*	0.060
	(0.252)	(0.168)	(0.199)	(0.182)	(0.177)	(0.152)	(0.108)	(0.199)
Age 35-44	-0.166	-0.152	-0.132	0.150	0.029	-0.033	-0.183*	-0.280
	(0.221)	(0.131)	(0.170)	(0.157)	(0.155)	(0.133)	(0.092)	(0.186)
Age 45-54	-0.443*	-0.263*	-0.316	0.093	-0.045	-0.113	-0.214*	-0.335
	(0.215)	(0.122)	(0.167)	(0.151)	(0.149)	(0.128)	(0.085)	(0.180)
Age 55-64	-0.405	-0.311**	-0.350*	-0.004	-0.154	-0.131	-0.327***	-0.466**
	(0.209)	(0.117)	(0.163)	(0.147)	(0.148)	(0.124)	(0.081)	(0.179)
Age 65+	-0.150	0.042	-0.331	0.127	-0.068	0.177	0.043	-0.278
	(0.281)	(0.215)	(0.306)	(0.218)	(0.216)	(0.269)	(0.166)	(0.220)
North-West	0.268	-0.087	0.195	0.178	-0.196	-0.171	0.002	-0.063
	(0.218)	(0.190)	(0.199)	(0.215)	(0.201)	(0.200)	(0.180)	(0.195)
Yorks/Humbs	0.024	0.262	0.191	0.314	0.149	-0.087	0.178	-0.013
	(0.228)	(0.199)	(0.201)	(0.219)	(0.204)	(0.207)	(0.190)	(0.193)
E. Midlands	0.165	-0.127	0.105	0.140	-0.142	-0.081	0.107	0.097
	(0.234)	(0.208)	(0.204)	(0.223)	(0.217)	(0.204)	(0.188)	(0.204)
W. Midlands	0.033	-0.020	0.043	0.126	0.062	0.010	0.093	0.155
	(0.222)	(0.206)	(0.205)	(0.219)	(0.206)	(0.202)	(0.181)	(0.186)
E. England	0.169	0.036	0.154	0.331	0.170	0.027	0.216	0.209
	(0.195)	(0.181)	(0.191)	(0.209)	(0.196)	(0.196)	(0.179)	(0.184)
London	0.100	-0.007	-0.117	0.072	-0.013	-0.191	0.033	0.048
	(0.200)	(0.180)	(0.186)	(0.206)	(0.192)	(0.191)	(0.173)	(0.171)
South-East	0.259	0.133	0.088	0.349	0.087	0.023	0.182	0.155
	(0.196)	(0.176)	(0.185)	(0.203)	(0.189)	(0.189)	(0.175)	(0.173)
South-West	0.011	-0.025	0.019	0.164	-0.157	-0.128	0.031	0.090
	(0.199)	(0.182)	(0.186)	(0.207)	(0.193)	(0.194)	(0.176)	(0.179)
Wales	-0.193	-0.169	-0.120	-0.109	-0.344	-0.317	-0.222	-0.063
	(0.251)	(0.204)	(0.215)	(0.230)	(0.228)	(0.226)	(0.232)	(0.232)
Scotland	0.083	0.092	0.176	0.286	0.085	-0.010	0.199	0.300
	(0.233)	(0.192)	(0.199)	(0.214)	(0.202)	(0.206)	(0.197)	(0.196)
N.Ireland	-0.041	0.597*	0.330	0.240	-0.028	-0.159	0.353	0.106
	(0.317)	(0.271)	(0.244)	(0.252)	(0.264)	(0.246)	(0.240)	(0.237)
N	1019	1325	1399	1522	1745	1664	1826	1856
pseudo R-sq	0.004	0.003	0.003	0.003	0.003	0.002	0.002	0.004

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Details of the cut points are available on request

Table 12: The Ordered Probit Models used to fit Investment Income in the SPI: Women, Labour Income not available

	2003	2004	2005	2006	2007	2009	2010	2013
oinvtota								
Age 25-34	-0.324	-0.348	-0.397	-0.106	-0.229	-0.482*	-0.190	-0.032
	(0.276)	(0.204)	(0.218)	(0.202)	(0.239)	(0.219)	(0.113)	(0.189)
Age 35-44	-0.202	-0.138	-0.350	0.149	-0.057	-0.449*	-0.099	-0.180
	(0.248)	(0.159)	(0.184)	(0.169)	(0.205)	(0.188)	(0.100)	(0.155)
Age 45-54	-0.173	-0.124	-0.374*	0.092	-0.138	-0.362*	-0.026	-0.159
	(0.242)	(0.153)	(0.182)	(0.165)	(0.201)	(0.181)	(0.092)	(0.144)
Age 55-64	-0.239	-0.213	-0.409*	0.010	-0.207	-0.475**	-0.109	-0.195
	(0.242)	(0.150)	(0.180)	(0.163)	(0.200)	(0.181)	(0.092)	(0.144)
Age 65+	0.412	0.135	-0.197	0.233	0.020	-0.379	0.056	-0.016
	(0.259)	(0.186)	(0.251)	(0.200)	(0.245)	(0.268)	(0.200)	(0.201)
North West	0.041	-0.350*	-0.336	-0.095	-0.090	-0.189	0.024	-0.013
	(0.212)	(0.169)	(0.178)	(0.192)	(0.186)	(0.195)	(0.190)	(0.213)
Yorks/Humbs	-0.178	-0.451*	-0.569**	-0.426*	-0.194	-0.148	-0.083	-0.182
	(0.218)	(0.177)	(0.188)	(0.195)	(0.196)	(0.194)	(0.199)	(0.221)
E. Midlands	-0.038	-0.448*	-0.441*	-0.364	-0.041	-0.005	-0.018	-0.215
	(0.220)	(0.175)	(0.189)	(0.200)	(0.200)	(0.202)	(0.193)	(0.221)
W. Midlands	0.223	-0.380*	-0.348	-0.240	-0.096	-0.081	-0.096	-0.203
	(0.219)	(0.177)	(0.188)	(0.191)	(0.192)	(0.192)	(0.192)	(0.224)
E. England	0.017	-0.302	-0.334	-0.233	-0.094	0.107	0.110	-0.026
	(0.205)	(0.163)	(0.171)	(0.184)	(0.177)	(0.179)	(0.180)	(0.206)
London	-0.098	-0.424*	-0.330	-0.254	-0.177	-0.100	-0.038	-0.219
	(0.205)	(0.165)	(0.172)	(0.181)	(0.175)	(0.175)	(0.177)	(0.199)
South-East	0.060	-0.233	-0.259	-0.119	0.033	0.052	0.168	-0.032
	(0.199)	(0.154)	(0.165)	(0.175)	(0.171)	(0.172)	(0.174)	(0.199)
South-West	-0.071	-0.396*	-0.403*	-0.322	-0.089	-0.112	-0.121	-0.161
	(0.203)	(0.161)	(0.171)	(0.179)	(0.177)	(0.178)	(0.182)	(0.202)
Wales	-0.162	-0.384	-0.402	-0.302	-0.233	-0.247	-0.059	-0.233
	(0.280)	(0.207)	(0.208)	(0.210)	(0.209)	(0.215)	(0.199)	(0.227)
Scotland	0.161	-0.115	-0.187	-0.050	0.086	0.227	0.219	0.071
	(0.223)	(0.171)	(0.180)	(0.190)	(0.185)	(0.191)	(0.194)	(0.217)
N.Ireland	-0.613	-0.757*	-0.440	-0.044	-0.209	-0.078	-0.076	-0.164
	(0.416)	(0.309)	(0.278)	(0.279)	(0.253)	(0.226)	(0.219)	(0.257)
N	1367	1724	1823	1876	1991	1700	1802	1878
pseudo R-sq	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.001

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Details of the cut points are available on request

Table 13: ML estimation of Pareto distribution for top of investment income

Sex	0.291***
	(0.013)
Financial year	-0.010***
	(0.002)
Constant	20.950***
	(3.153)
N	108789

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B The Distribution of Pension Rights and Investment Income

Table 14: The Distribution of Pension and Insurance Wealth, 2012-2014

Upper Limit (£)	Empl Inc	No Empl Inc	Pension Inc	No Pens Inc	Total	Proportion
0	2,543,535	1,888,556	-	1,633,342	6,065,433	23.69%
20000	2,544,062	464,274	794,214	142,597	3,945,147	15.41%
40000	1,287,470	178,273	533,176	53,825	2,052,744	8.02%
60000	932,775	124,978	390,211	18,533	1,466,497	5.73%
80000	788,446	81,751	368,433	20,631	1,259,261	4.92%
100000	664,849	82,774	298,447	12,892	1,058,961	4.14%
120000	527,611	68,726	272,975	11,219	880,530	3.44%
140000	355,759	46,485	234,998	8,108	645,350	2.52%
160000	356,319	43,740	225,600	7,029	632,689	2.47%
180000	352,486	30,821	198,404	8,091	589,803	2.30%
200000	315,340	46,555	153,181	12,873	527,949	2.06%
225000	320,919	38,403	187,289	9,942	556,554	2.17%
250000	292,553	35,957	132,765	956	462,231	1.80%
300000	525,707	42,258	256,829	16,813	841,607	3.29%
350000	398,491	46,297	246,424	5,769	696,981	2.72%
400000	262,558	39,383	174,669	5,263	481,873	1.88%
450000	256,920	38,947	139,034	7,132	442,033	1.73%
500000	210,570	30,134	146,921	1,064	388,688	1.52%
600000	370,078	66,700	164,352	772	601,901	2.35%
700000	281,005	45,884	133,878	6,553	467,320	1.82%
800000	189,212	29,360	90,477	993	310,042	1.21%
1000000	276,057	56,021	128,838	2,497	463,413	1.81%
1250000	207,372	49,317	65,858	4,009	326,556	1.28%
1500000	101,238	31,587	47,295	2,624	182,744	0.71%
1750000	61,068	19,763	24,193		105,024	0.41%
2000000	40,508	4,267	11,946	609	57,330	0.22%
2500000	29,781	4,908	16,836	501	52,026	0.20%
3000000	2,878	3,954	2,269		9,102	0.04%
3500000	9,826	1,135	2,637		13,597	0.05%
4000000	6,161	926	1,063		8,150	0.03%
4500000	487	1,478	2,990		4,955	0.02%
5000000	2,406	339			2,744	0.01%
5.00E+07	3,211	2,783	3,202		9,196	0.04%

Table 15: Distribution of investment income, 2013

oinvtot	Sum	Mean	Count	Prop (Count)	Prop (Sum)
1	0	0	5,258,200	17.46	-
101	365,266,383	23	15,969,746	53.04	0.57
201	253,993,759	149	1,702,420	5.65	0.40
301	158,564,095	249	637,007	2.12	0.25
401	176,313,961	352	501,540	1.67	0.28
501	303,256,639	459	660,992	2.20	0.47
601	518,963,879	560	926,435	3.08	0.81
701	235,039,838	639	367,853	1.22	0.37
801	292,083,216	736	396,629	1.32	0.46
901	185,653,617	860	215,962	0.72	0.29
1001	147,005,984	951	154,659	0.51	0.23
1501	588,108,828	1,228	478,861	1.59	0.92
2001	577,684,431	1,747	330,590	1.10	0.90
2501	399,833,827	2,227	179,507	0.60	0.63
3001	327,965,582	2,752	119,170	0.40	0.51
3501	349,218,456	3,245	107,625	0.36	0.55
4001	316,542,669	3,756	84,269	0.28	0.50
4501	319,141,068	4,255	75,002	0.25	0.50
5001	358,513,457	4,741	75,617	0.25	0.56
5501	270,958,534	5,248	51,628	0.17	0.42
6001	373,257,093	5,711	65,358	0.22	0.58
6501	306,435,515	6,240	49,106	0.16	0.48
7001	368,112,394	6,741	54,604	0.18	0.58
7501	293,813,225	7,255	40,496	0.13	0.46
8001	315,010,972	7,753	40,633	0.13	0.49
8501	333,143,062	8,261	40,329	0.13	0.52
9001	356,699,548	8,778	40,634	0.13	0.56
9501	318,884,990	9,250	34,475	0.11	0.50
10001	354,452,867	9,799	36,173	0.12	0.55
10501	303,777,882	10,257	29,617	0.10	0.48
11001	286,925,805	10,752	26,685	0.09	0.45
11501	546,461,323	11,193	48,824	0.16	0.86
12001	282,442,250	11,762	24,014	0.08	0.44
12501	309,196,167	12,268	25,203	0.08	0.48
13001	261,710,157	12,761	20,509	0.07	0.41
13501	399,914,379	13,284	30,105	0.10	0.63
14001	320,728,695	13,790	23,259	0.08	0.50
14501	341,702,137	14,292	23,908	0.08	0.53
15001	316,493,336	14,785	21,406	0.07	0.50
15501	225,874,756	15,248	14,814	0.05	0.35
16001	362,592,281	15,733	23,046	0.08	0.57
16501	301,075,739	16,259	18,517	0.06	0.47
17001	589,951,547	16,741	35,239	0.12	0.92
17501	294,339,540	17,280	17,033	0.06	0.46
18001	324,669,317	17,786	18,254	0.06	0.51
18501	307,104,555	18,260	16,819	0.06	0.48
19001	335,358,829	18,797	17,841	0.06	0.52
19501	313,716,222	19,286	16,266	0.05	0.49
20001	380,341,403	19,855	19,156	0.06	0.60
20501	264,900,520	20,235	13,091	0.04	0.41
21001	299,184,893	20,759	14,412	0.05	0.47
21501	341,184,310	21,227	16,073	0.05	0.53
22001	296,303,903	21,769	13,611	0.05	0.46
22501	819,037,725	22,245	36,819	0.12	1.28
23001	333,111,837	22,762	14,635	0.05	0.52
23501	333,884,498	23,267	14,350	0.05	0.52
24001	280,262,230	23,781	11,785	0.04	0.44
24501	394,660,218	24,309	16,235	0.05	0.62
25001	364,945,539	24,822	14,702	0.05	0.57
25501	292,371,067	25,265	11,572	0.04	0.46
26001	389,106,410	25,739	15,118	0.05	0.61
26501	311,688,031	26,254	11,872	0.04	0.49
27001	533,912,457	26,759	19,952	0.07	0.84
27501	345,873,825	27,264	12,686	0.04	0.54
28001	795,272,017	27,803	28,604	0.10	1.24
28501	380,253,362	28,258	13,456	0.04	0.60
29001	436,750,722	28,806	15,162	0.05	0.68
29501	396,103,466	29,299	13,520	0.04	0.62
30001	466,585,164	29,848	15,632	0.05	0.73
30501	348,884,275	30,246	11,535	0.04	0.55
31001	453,708,780	30,735	14,762	0.05	0.71
31501	646,928,821	31,203	20,733	0.07	1.01
32001	564,159,343	31,788	17,747	0.06	0.88
32501	592,798,830	32,245	18,384	0.06	0.93
33001	484,466,848	32,801	14,770	0.05	0.76
33501	1,826,738,757	33,306	54,846	0.18	2.86
34001	1,072,877,088	33,750	31,789	0.11	1.68
34501	506,905,047	34,273	14,790	0.05	0.79
35001	398,267,564	34,797	11,446	0.04	0.62
35501	281,642,175	35,240	7,992	0.03	0.44
36001	418,596,132	35,707	11,723	0.04	0.66
36501	319,678,325	36,234	8,823	0.03	0.50
37001	354,204,512	36,752	9,638	0.03	0.55
37501	290,132,843	37,260	7,787	0.03	0.45
38001	301,212,374	37,778	7,973	0.03	0.47
38501	227,899,627	38,239	5,960	0.02	0.36
39001	377,313,967	38,859	9,710	0.03	0.59
39501	233,732,676	39,270	5,952	0.02	0.37
40001	310,642,776	39,853	7,795	0.03	0.49
40501	29,837,887,899	91,935	324,554	1.08	46.70

Note: Authors' calculations from SPI (2013).