

Technical Paper

IDS Under-Coverage

March 2018

(Revised for publication September 2019)

Strategy, Evaluation & Reporting Branch

Technical Paper IDS Under-Coverage

The authors are Philip O'Rourke and Jackie Mahon, Strategy, Evaluation & Reporting Branch, of the Office of the Revenue Commissioners (porourke@revenue.ie, jmahon@revenue.ie). Any opinions expressed in this paper are the views of the authors and do not necessarily reflect the views of the Office of the Revenue Commissioners.

Executive Summary

Revenue's Income Distribution Statistics (IDS) database provides data for detailed statistical analyses on PAYE and self-assessment Income Tax taxpayers. IDS is used extensively in Revenue's tax modelling work and in preparation of published statistics.

IDS is primarily based on tax returns, P35 returns for PAYE employers¹ due by February in the following tax year and Form 11 returns for self-assessed Income Tax cases due in November of the following year. Tax returns continue to be received or amended, in small numbers, for months and even years after the due date.

For the purposes of compiling and publishing statistics on tax returns, the IDS database is created in April of each year (for example, April 2017 saw the creation of IDS for tax year 2015 as returns for that year were filed through 2016). This is to facilitate the need for fixed and timely statistics on taxpayers required by stakeholders each year.

Any subsequent tax returns filed, or amendments to existing returns, are not incorporated in IDS. Instead, numbers and costs relating to data sourced from the IDS are grossed up to account for under-coverage in the IDS due to returns not being received at the time when the IDS is created.

Revenue used the following coverage percentages for the latest IDS, tax year 2015.

| | PAYE | Schedule D |
|-------------------------------|------|------------|
| Numbers & Cost/Tax | 96% | 85% |

As part of a review of the quality of statistical outputs, Revenue has undertaken an analysis of the IDS under-coverage. Based on data across six years of returns, it is proposed to have four growth percentages to replace the current method of applying two growth rates, set out as shown in the Table below.

| | Schedule D | Schedule E |
|-----------------|------------|------------|
| Numbers | 88.0% | 96.5% |
| Cost/Tax | 94.5% | 96.5% |

Revenue will commence using these revised percentages for under-coverage for statistics on tax year 2016. These will be reviewed periodically again in the future to ensure they remain valid.

¹ Prior to PAYE Modernisation, effective from 1 January 2019.

1 Introduction

The Income Distribution Statistics (IDS) database provides data for detailed reporting and statistical analyses on PAYE and self-assessment data and is used extensively in Revenue.

Tax returns continue to be received months and years after the due date of a filing period. However, for the purpose of compiling and publishing statistics on tax returns, the IDS is created in April of each year and any subsequent tax returns filed, or amendments to existing returns, are not incorporated. This is due to the need for fixed and timely statistics on taxpayers required by stakeholders each year.

Numbers and costs taken from the IDS are grossed up in the 'Cost of Tax Expenditures' publication² as well as in Revenue's Tax Modeller³ platform to account for under coverage in the IDS due to returns not being received at the time when the IDS is created. It is important to ensure that the estimated coverage of tax returns in any particular year, at the point of creating the IDS, is as accurate as possible, since any inaccuracies can lead to over-estimating or under-estimating the cost of any particular expenditure, or the cost or yield associated with a change in income tax policy.

² <https://www.revenue.ie/en/corporate/information-about-revenue/statistics/tax-expenditures/costs-expenditures.aspx>.

³ In order to provide costings for changes to Income Tax policy, for example for changes to tax rates or bands, we use Revenue's microsimulation model called Tax Modeller. This model is the main tool used in providing costings on income tax in each Budget cycle. The model is based on the latest actual data/returns available. We then grow this base up to account for population changes, income growth and other macros and elasticities, which come from the Department of Finance. Once this base is grown up it then represents the estimated tax base for the next tax year.

2 Outline of Schedule D and Schedule E Taxpayers

For analysis purposes in Revenue, we classify taxpayers as being either Schedule E/PAYE or Schedule D/ Self-Assessed, depending on their source of income and the relative value of that income, as set out below.

1. Schedule D/Self-Assessed

These are taxpayers who have earned or unearned income, which do not have tax automatically collected by an employer on behalf of Revenue through the PAYE system. These cases file a Form 11 return through ROS or through paper returns to Revenue. Prior to this review it was estimated that approx. 85% of all Form 11s for a particular year are received and processed when the file is extracted to create the IDS.

2. PAYE/Schedule E

PAYE employment details are gathered through P35 employer returns⁴ and Form 12 returns. Prior to this review it was estimated that approx. 96% of P35/Form 12 returns are filed with Revenue when the file is extracted to create the IDS. It is important to bear in mind that, for the purpose of the IDS, the PAYE returns are compiled around February of each year, a year after the original deadline for receipt of these returns. This is due to the timing of the filing of Form 11s, which are due to be filed 10/11 months after the end of the tax year. For example, PAYE details for 2016 are received in February of 2017 and Form 11 details for 2016 are received by October/November of 2017. The PAYE file that is then used in the 2016 IDS is compiled in February 2018 and has already 'aged' for a period of 12 months, by which it is meant that since 12 months have passed it is more likely that any late filers will have made their return, or any amendments that need to be made to the original returns will have been completed. This is in comparison to the Form 11 returns, which have only 'aged' for less than 5 months.

Timeline for 2016 Returns:

- Tax year runs from 01/01/2016 to 31/12/2016
- Employer P35 returns due February 2017
- Self-assessed Form 11 returns due by November 2017
- Employer returns (PAYE) for the IDS processed in February 2018
- Accessing file (containing Form 11 data) finalised and IDS created early April 2018

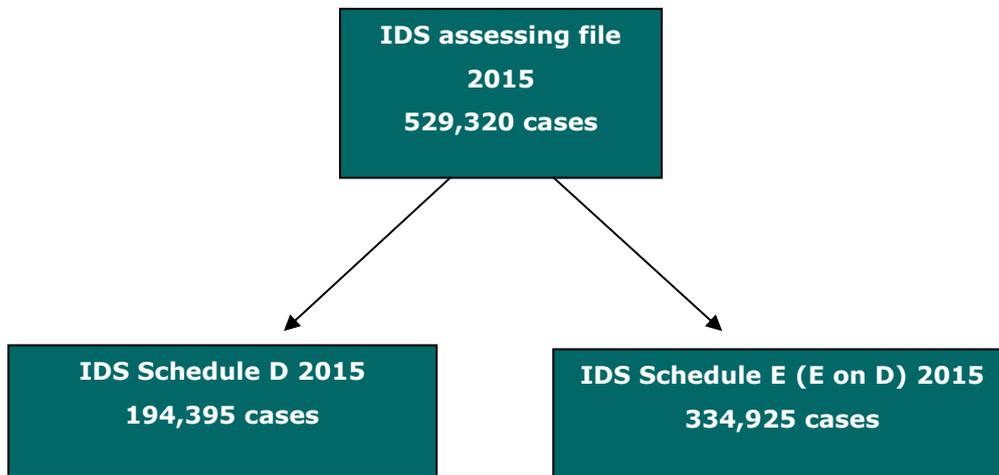
3. Schedule D and Schedule E on D

Those taxpayers who file a Form 11 may also have a PAYE source of income. The Assessing file is split into 2 categories, Schedule D and Schedule E on D. E on D cases are PAYE cases in the self-assessed system, that is, PAYE taxpayers who also have a source of non-PAYE income and have completed a Form 11 return in order to declare the non-PAYE

⁴ Prior to PAYE Modernisation, effective from 1 January 2019.

portion of their income. The determination of category that a taxpayer unit falls in to is based on the following criteria:

- If non-PAYE income > 50% of gross income, then the tax unit is designated as a Schedule D case.
- If PAYE income > 50% of gross income, then the tax unit is designated as a Schedule E case.



The diagram above shows the split between Schedule D and Schedule E on D cases in 2015. These Schedule E on D cases are then combined with the PAYE/Schedule E cases to create the total Schedule E case-base. As shown, most Form 11 returns relate to tax payers who have a PAYE source of income and who have non-PAYE income of less than 50% of their gross income. Schedule D cases only account for 8.5% of the total amount of cases on the All Persons file (Schedule D plus Schedule E cases), which was 2,301,798 taxpayer units in 2015. Only about 37% of Form 11 returns relate to taxpayers that we classify as Schedule D taxpayers and of these about 30% also have a PAYE income (of less than 50% of their gross income).

The Schedule E cases are currently assigned the 96% coverage in the IDS at the point of creation of the IDS, while the remaining 194,395 Schedule D cases are currently assigned the 85% coverage.

3 Coverage Percentage – Form 11 Returns

The following table shows the coverage percentages which were used when grossing up Schedule D figures each tax year since 2011.

| 2011 | 2012 | 2013 | 2014 | 2015 |
|-------|-------|-------|-------|-------|
| 83.0% | 83.0% | 82.7% | 85.9% | 84.9% |

These percentages were based on information in relation to the number of Form 11s issued and represented the percentage of coverage if all Form 11s that were issued would eventually be returned. However, it may be the case that this figure, which represents the target of 100% return rate, never materialises, as the number of forms issued/businesses registered may not be up to date and reflective of those tax payers required to make a return.

An alternative approach is to look at the number of Form 11s received after a longer period of time has passed such that the file has matured for a longer period of time, allowing for receipt of late returns in relation to a tax year. The tables below compare the number of Form 11 returns on Revenue's 'Live' system on 22/02/2018 with the number of Form 11 returns on our IDS Assessing file for each year 2011 – 2016.

| | Current Form 11 data (live files) | | | IDS Form 11 data (static files) | | |
|------|-----------------------------------|--------------------------------|-------------------|---------------------------------|--------------------------------|-------------------|
| | Number of Form 11 Returns | Form 11 Returns with Tax Due>0 | | Number of Form 11 Returns | Form 11 Returns with Tax Due>0 | |
| | | Number | Sum of Tax Due €m | | Number | Sum of Tax Due €m |
| 2011 | 526,405 | N/A | N/A | 459,404 | 318,395 | 4,654.00 |
| 2012 | 537,361 | N/A | N/A | 473,550 | 332,129 | 4,976.03 |
| 2013 | 546,956 | 381,073 | 5,533.66 | 488,100 | 349,570 | 5,227.38 |
| 2014 | 560,416 | 402,464 | 5,989.37 | 519,738 | 378,972 | 5,755.86 |
| 2015 | 567,395 | 419,540 | 6,869.06 | 529,320 | 397,236 | 6,656.68 |
| 2016 | 511,818 | 385,415 | 6,141.41 | 503,601 | 380,908 | 6,089.49 |

| | Form 11 Returns with Tax Due>0 | | | | | |
|------|---|--|---|--|--|--|
| | Difference between Live file and Static file in Number of Tax Returns | Number of Returns in Static file as a % of the Live file | Difference between Live file and Static file in Number of Tax Returns | Number of Returns in Static file as a % of the Live file | Difference between Live file and Static file in Amount of Tax Due €m | Amount of Tax Due in Static file as a % of the Live file |
| 2011 | 67,001 | 87.3% | | | | |
| 2012 | 63,811 | 88.1% | | | | |
| 2013 | 58,856 | 89.2% | 31,503 | 91.7% | 306.28 | 94.5% |
| 2014 | 40,678 | 92.7% | 23,492 | 94.2% | 233.51 | 96.1% |
| 2015 | 38,075 | 93.3% | 22,304 | 94.7% | 212.38 | 96.9% |
| 2016 | 8,217 | 98.4% | 4,507 | 98.8% | 51.92 | 99.2% |

The following table sets out these findings alongside the actual historical under-coverage percentages used in previous years.

| | Number of Returns in Static file as a % of the Live file | Amount of Tax Due in Static file as a % of the Live file | Historical under coverage % Used for both Numbers and Tax Due |
|------|--|--|---|
| 2011 | 87.3% | | 83.0% |
| 2012 | 88.1% | | 83.0% |
| 2013 | 89.2% | 94.5% | 82.7% |
| 2014 | 92.7% | 96.1% | 85.9% |
| 2015 | 93.3% | 96.9% | 84.9% |

From this analysis it appears that the under-coverage has been overestimated in the period 2011 to 2015. In 2011 it was estimated that the assessing file in our IDS was 83% complete, however looking at the table above we see that our IDS file is 87.3% complete when compared with the number of Form 11 returns received for 2011 by February 2018, 5 years and 3 months after the deadline for filing a Form 11 return for 2011.

Looking to the number of returns for 2014, our IDS file is 92.7% complete when compared with the number of Form 11 returns for 2014 received by February 2018, and similarly the equivalent figure for 2015 returns is 93.3%. This may be due to the fact that further time is needed for these files to mature sufficiently. For example, returns for 2015 were due by November 2016, therefore only 1 year and 3 months had passed by the time the data was collated. Likewise, only 2 years and 3 months have passed since the 2014 returns were due.

An analysis of returns received in the most recent 3 years for the tax year 2012 shows that in the order of 40% of all late returns (late being defined in this context as being filed after the formation of the IDS for that tax year) were received in this period, meaning that with the further passing of time it is likely the number of returns for the most recent years will grow further, hence it is clear that the coverage percentages for the later years in the table above are not suitable for use, since they overestimate the coverage percentage. For this reason, it is clear that the coverage percentages to focus on are those for the years 2011, 2012 and 2013.

Looking to the returns which have a positive tax due amount in 2014, the IDS file is 94.2% complete for number of returns and 96.1% complete for the amount of tax due compared to the live data for those years (February 2018). The reason for these higher percentages when looking at returns where tax due is greater than zero, as opposed to all returns, may be due to the fact that late filers often have no tax liability, and therefore are not liable for a penalty due to late filing.⁵ For the purpose of this review it was decided to

⁵ Surcharge for Late Returns You must return the completed 2016 Tax Return on or before 31 October 2017. If your Return is late a 'late-filing' surcharge will apply. This surcharge, which, is added on to your tax due, is:

- 5% of the tax due or €12,695, whichever is the lesser, where the Return is submitted after 31 October 2017 and on or before 31 December 2017,
- 10% of the tax due or €63,485, whichever is the lesser, where the Return is submitted after 31 December 2017.

focus on all tax returns, rather than only looking returns where there was tax due, but it is clear that it is necessary to use separate growth figures for number of tax returns filed and tax/cost, to reflect the fact that the under-coverage of the IDS differs when referring to numbers of returns and the total tax received.

In relation to the coverage percentage for cost/tax data from Form 11s, it is apparent that the current practice of assuming a coverage percentage of 85% needs to be revised. The table above shows that in 2013, 94.5% of the tax due was included in the IDS file, compared to the current Live data. As this is the oldest year for which it was possible to collate this data, allowing for the most time for the Live file to mature further, this was chosen as the appropriate coverage percentage to apply for under-coverage for cost/tax data. The coverage percentage for numbers of taxpayers in the IDS compared to the Live file is 89% in 2014, which is close to the average for the years 2011-2013 (88%), which gives some confidence in taking the 2014 figure of 94.5% as the appropriate Schedule D coverage percentage for cost/tax data.

4 Coverage Percentage – PAYE

The PAYE returns used in the IDS for each of the years 2011-2015 were assumed to represent 96% coverage of the entire PAYE returns. The 96% coverage was applied to each year, as the P35 data are assumed to be almost complete when it is extracted and loaded into the IDS – as referenced previously, the PAYE file has aged for a longer period of time compared to the Form 11 returns. The P35 is an annual return that is completed by all registered employers after the tax-year end. As PAYE returns are the biggest component of Income Tax revenue the compliance and enforcement of PAYE returns is generally high.

All P35s for 2015 should be returned by February 2016. As the data was not extracted for our 2015 PAYE IDS files until February/March 2017, it is clear that the majority of the PAYE files have been captured in the IDS.

Due to the way in which the PAYE file is compiled for the IDS each year it is not possible to re-create this file in order to assess how the figures have changed since these files were originally created. However, some comparisons/validations are possible using alternative sources in Revenue. The following table compare the PAYE earnings on the P35DTL files in Revenue's live system with the PAYE total earnings on the IDS files for each year 2011 – 2016.

| | IDS file | | Live System files | | Coverage |
|------|---------------------|-----------|-------------------|-----------|---------------|
| | PAYE Total Earnings | | Total Amount Pay | | |
| | Number | Sum €m | Number | Sum €m | PAYE Earnings |
| 2011 | 1,888,269 | 61,605.45 | 2,274,446 | 64,526.77 | 95.5% |
| 2012 | 1,940,945 | 63,554.67 | 2,287,983 | 65,024.32 | 97.7% |
| 2013 | 1,976,852 | 64,805.63 | 2,336,642 | 66,432.70 | 97.6% |
| 2014 | 2,044,033 | 67,624.82 | 2,415,959 | 69,246.27 | 97.7% |
| 2015 | 2,120,409 | 71,769.96 | 2,509,090 | 73,530.27 | 97.6% |
| 2016 | 2,307,296 | 75,730.96 | 2,612,482 | 77,928.56 | 97.2% |

It is not possible to compare the numbers in the table above as the numbers on the IDS All Persons files are taxpayer units whereas the numbers on the P35DTL files are individuals. It is also not possible to compare the 'Tax_Due' figures as 'Tax_due' in the IDS All Persons file contains tax from all other incomes, not just PAYE income.

Instead, the PAYE total earnings figures were the only comparable figures that could be used on both files, and as can be seen above the coverage of our IDS PAYE data in relation to earnings is just under 98% for each of the years 2012-2015. This ties in well with our current methodology of accounting for under coverage, where we assume the figures in the IDS represent a coverage rate of 96%. Given that the current figure used is 96%, and that it is not possible to separately calculate the coverage for taxpayer numbers, it is proposed that a rate of 96.5% be applied to both costs/tax and numbers for PAYE cases.

5 Recommendations and Conclusion

Following an analysis of the tax records for previous years it is proposed to have four growth percentages to apply to figures from the IDS, to replace the current method of applying two growth rates, they are:

| | Schedule D | Schedule E |
|-----------------|------------|------------|
| Numbers | 88.0% | 96.5% |
| Cost/Tax | 94.5% | 96.5% |

It is proposed to apply these revised percentages for under coverage to the Cost of Tax Expenditures table, commencing with the table to be published in July 2018 that will contain data on credit and reliefs for the tax year 2016, as well as to the creation of the base for use in Tax Modeller - '2019base2016'.

The figure for the number of returns in our IDS file as a percentage of the number of returns in the Live file increased by 1% in each of the years 2011-2013 inclusive, from 87% in 2011 to 89% in 2013. It is possible there is further scope for increasing this percentage even further, given that the number of Form 11 returns made online through ROS also increased by about 1% in each of those years, from 83.1% in 2011 to 86% in 2013. As those figures have continued to increase since then, to 87.5% in 2014 and 89.9% in 2015, it is feasible that the increasing use of ROS may have had an upswing effect on the timeliness of returns. However, it was decided on this occasion to use an average of the coverage percentage for each of the years 2011 to 2013.

It is possible that the revised coverage figure of 94.5% for the cost/tax in Schedule D cases may lead to a decrease in some credits and reliefs for 2016, given that the growth in those figures between 2015 and 2016 may be less than the difference between applying a grossing up rate of 85% and 94.5% for the Schedule D component. However, given the analysis in this paper it seems clear that the costs of credits and reliefs were being somewhat over-estimated.

Finally, in relation to the PAYE/Schedule E cases, the average of the earnings coverage percentage between 2011 and 2013 (97%) was taken along with the current coverage percentage used for the number of cases in those years (96%) to give a final coverage figure of 96.5%. This is very close to the current percentage of coverage used and reflects the fact that the PAYE files are, by their nature, more complete in comparison to Form 11 returns.

This analysis will be reviewed periodically in future to ensure the approach continues to use the most up to date information available. This will be particularly important post the implementation and bedding in of PAYE Modernisation (from 1 January 2019) as this fundamentally alters the reporting on the Schedule E/PAYE components.