<table>
<thead>
<tr>
<th>Chapter A: Introduction</th>
<th>..........................................................</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methodology</td>
<td>......................................................................................................................</td>
<td>3</td>
</tr>
<tr>
<td>Methods used to calculate the tax gap</td>
<td>..................................................................................................................</td>
<td>4</td>
</tr>
<tr>
<td>Chapter B: Accuracy and reliability</td>
<td>...............................................</td>
<td>5</td>
</tr>
<tr>
<td>Value Added Tax</td>
<td>......................................................................................................................</td>
<td>5</td>
</tr>
<tr>
<td>Excise duties</td>
<td>......................................................................................................................</td>
<td>6</td>
</tr>
<tr>
<td>Direct taxes</td>
<td>......................................................................................................................</td>
<td>6</td>
</tr>
<tr>
<td>Chapter C: Tax gap and compliance yield</td>
<td>................................................</td>
<td>7</td>
</tr>
<tr>
<td>Chapter D: Value Added Tax</td>
<td>..................................................................................................................</td>
<td>9</td>
</tr>
<tr>
<td>VAT gap</td>
<td>......................................................................................................................</td>
<td>9</td>
</tr>
<tr>
<td>Chapter E: Alcohol</td>
<td>......................................................................................................................</td>
<td>12</td>
</tr>
<tr>
<td>Spirits and beer (upper bound) estimate</td>
<td>........................................................................................................</td>
<td>12</td>
</tr>
<tr>
<td>Beer lower estimate</td>
<td>......................................................................................................................</td>
<td>16</td>
</tr>
<tr>
<td>Wine central estimate</td>
<td>......................................................................................................................</td>
<td>21</td>
</tr>
<tr>
<td>Chapter F: Tobacco</td>
<td>......................................................................................................................</td>
<td>22</td>
</tr>
<tr>
<td>Overview</td>
<td>......................................................................................................................</td>
<td>22</td>
</tr>
<tr>
<td>Methodology</td>
<td>......................................................................................................................</td>
<td>22</td>
</tr>
<tr>
<td>Chapter G: Diesel</td>
<td>......................................................................................................................</td>
<td>28</td>
</tr>
<tr>
<td>Methodology</td>
<td>......................................................................................................................</td>
<td>28</td>
</tr>
<tr>
<td>Chapter H: Estimates using random enquiry programmes</td>
<td>.......................................</td>
<td>30</td>
</tr>
<tr>
<td>Random enquiry programme estimates</td>
<td>........................................................................................................</td>
<td>30</td>
</tr>
<tr>
<td>Populations and sampling</td>
<td>................................................................................................................</td>
<td>30</td>
</tr>
<tr>
<td>Data features</td>
<td>......................................................................................................................</td>
<td>33</td>
</tr>
<tr>
<td>Sources of error</td>
<td>......................................................................................................................</td>
<td>35</td>
</tr>
<tr>
<td>Modelling adjustments</td>
<td>......................................................................................................................</td>
<td>35</td>
</tr>
<tr>
<td>Validation</td>
<td>......................................................................................................................</td>
<td>36</td>
</tr>
<tr>
<td>Outliers</td>
<td>......................................................................................................................</td>
<td>36</td>
</tr>
<tr>
<td>Deselections</td>
<td>......................................................................................................................</td>
<td>36</td>
</tr>
<tr>
<td>Enquiry data</td>
<td>......................................................................................................................</td>
<td>36</td>
</tr>
<tr>
<td>Tax gap calculation</td>
<td>......................................................................................................................</td>
<td>37</td>
</tr>
<tr>
<td>Non-payment</td>
<td>......................................................................................................................</td>
<td>38</td>
</tr>
<tr>
<td>Compliance yield</td>
<td>......................................................................................................................</td>
<td>38</td>
</tr>
<tr>
<td>Other estimates</td>
<td>......................................................................................................................</td>
<td>38</td>
</tr>
<tr>
<td>Chapter I: Estimates using risk-based enquiry programmes</td>
<td>..................................</td>
<td>40</td>
</tr>
<tr>
<td>Risk-based estimates</td>
<td>......................................................................................................................</td>
<td>40</td>
</tr>
<tr>
<td>Extreme value methodology</td>
<td>..........................................................................................................</td>
<td>40</td>
</tr>
<tr>
<td>Sources of error</td>
<td>......................................................................................................................</td>
<td>42</td>
</tr>
<tr>
<td>Non-payment</td>
<td>......................................................................................................................</td>
<td>43</td>
</tr>
<tr>
<td>Compliance yield</td>
<td>......................................................................................................................</td>
<td>43</td>
</tr>
<tr>
<td>Chapter J: Other taxes</td>
<td>......................................................................................................................</td>
<td>44</td>
</tr>
<tr>
<td>Stamp Duty Land Tax</td>
<td>......................................................................................................................</td>
<td>44</td>
</tr>
<tr>
<td>Landfill Tax</td>
<td>......................................................................................................................</td>
<td>46</td>
</tr>
<tr>
<td>Chapter K: Avoidance and hidden economy</td>
<td>........................................</td>
<td>48</td>
</tr>
<tr>
<td>Avoidance</td>
<td>......................................................................................................................</td>
<td>48</td>
</tr>
</tbody>
</table>
Hidden economy ...................................................................................................................................... 49
Chapter L: Abbreviations ....................................................................................................................... 52
Chapter A: Introduction

Methodology

A.1 This document provides further details of the data and methodology used to produce estimates of the tax gap published in ‘Measuring tax gaps 2020 edition’.

A.2 There are numerous methodological approaches to measuring tax gaps.

A.3 Top-down methods use external independent data sources to estimate total consumption of taxable products to calculate the total theoretical liabilities; the tax gap is the difference between the total theoretical liabilities and the tax actually paid. An example of this is the Value added Tax (VAT) gap.

A.4 Bottom-up methods include a number of techniques:

- random enquiry programs – these are full enquiries opened by HMRC compliance officers into a randomly selected sample of taxpayers
- statistical methods – unlike random enquiry programs these use risk-based enquiries that are not representative of the whole population, and require statistical methods to scale up the results to the whole population
- management information – these methods use management information such as:
  o risk registers (a list of identified tax risks, together with information such as estimated value, nature and status)
  o data extracted from accounting systems
  o other databases or systems used to manage HMRC’s business

A.5 Experimental methodologies are used to produce illustrative estimates where there is no direct measurement data. For these tax gap components, we use the best available data, simple models and are produced using assumptions made in collaboration with HMRC’s operational experts to build an illustrative estimate of the tax gap.

A.6 We employ the most appropriate methodology for each tax gap component, based on the factors listed below:

- availability of quality HMRC data
- availability of quality independent data
- structure of the tax regime
- cost and impact for both HMRC and taxpayers
- level of granularity demanded

A.7 Generally, following good international practices, we use ‘top-down’ methodology for indirect taxes and ‘bottom-up’ methodologies for direct taxes. The tax gap estimates may, however, also be produced by compiling the results from a combination of two or more methods.
Methods used to calculate the tax gap

A.8 Figure A.1 below shows the general methodological approach used to calculate each tax gap component.

Figure A.1 Tax gap methodologies

A.9 Figure A.2 below shows a summary of the tax gap by methodology. A degree of assumption and judgement has been applied to attribute some elements of the tax gap to methodology types, especially where a combination of methods is used.

Figure A.2 Tax gap by methodology
Chapter B: Accuracy and reliability

B.1 Our tax gap estimates are official statistics produced with the highest levels of quality assurance and adhere to the framework for the Code of Practice for Statistics, guidance for which can be found at [www.statisticsauthority.gov.uk/code-of-practice/](http://www.statisticsauthority.gov.uk/code-of-practice/).

B.2 This code assures objectivity and integrity – providing the framework to ensure that statistics are trustworthy, good quality, and are valuable and provides producers of official statistics with the detailed practices they must commit to when producing and releasing official statistics.

B.3 The figures presented in ‘Measuring tax gaps 2020’ edition are our best estimates based on the information available, but there are sources of uncertainty and potential error. This means that the estimate might differ from the “true value” so it is best to focus on the trend in the results rather than the absolute numbers when interpreting the findings.

B.4 Statistical uncertainty is caused by two factors:

- sampling error – errors that arise because the estimates rely on information collected from a sample, rather than from the whole population; sampling error can lead to year-on-year fluctuations in the tax gap estimates that do not reflect true changes in the size of the tax gap
- bias or non-sampling error – systematic errors where the modelling assumptions or errors in the data lead to estimates that are consistently either too low or too high

B.5 Where possible, HMRC has estimated the likely impact of sampling errors by calculating statistical confidence intervals. These give margins of error within which we would expect the true value lies 95% of the time, if there were no systematic errors. They provide an indication of the extent to which changes in the estimates between years can be confidently interpreted as true changes. They do not take account of systematic errors that might lead the central estimate to be too low or too high over the whole series.

B.6 Systematic error is less straightforward to deal with, as it is not defined by statistical assessments that allow for easy interpretation. In order to give an indication of the effect of these biases HMRC presents the tax gaps for alcohol and tobacco as ranges. For beer and tobacco these are constructed as the range between upper and lower bounds, representing the degree of uncertainty associated with those systematic biases for which upper and lower bounds can be derived.

Value Added Tax

B.7 The VAT Total Theoretical Liability (VTTL) model and the top-down VAT gap derived from it are broad measures, subject to a degree of uncertainty. They are based on an analysis of survey and other data and include a number of assumptions and adjustments which add both random and systematic variation to the estimates. There is also a small element of forecasting in some of the spending data, which introduces further variation.

B.8 It is not possible to produce a precise confidence interval for the VAT revenue loss estimates. The VTTL estimate is constructed largely from Office for National Statistics (ONS) National Accounts data which is derived, in the main, from sample surveys and is thus subject to both sampling and non-sampling errors. The ONS does not publish error margins for the relevant input series and so it is not possible to construct an estimate of the impact of these errors on the VTTL.

B.9 The VAT gap is updated and revised as and when new data becomes available, or new methodologies are developed. HMRC publishes a revised historical VAT gap series once a year in the ‘Measuring tax gaps’ publication, incorporating both new and updated data and methodological improvements together. The VAT gap preliminary estimate for tax year 2019 to 2020 is expected to be published on the day of Autumn Budget 2020 and a second estimate is expected to be
Excise duties

Systematic biases

B.10 Systematic biases are explicitly considered for beer and tobacco products, with results presented as a range to represent the degree of uncertainty. These ranges are discussed in Chapter E for beer and Chapter F for tobacco products.

B.11 No account is presently made for systematic biases in the spirits and diesel estimates.

Random variation

B.12 While the upper and lower estimates for beer and tobacco will contain random variation, the resulting confidence intervals are not shown in this document as these estimates are used to represent the uncertainty around our central estimate.

B.13 For spirits, an assessment of the effect of random variation is included using error margins. These are estimated by combining the random errors (where available) from all data sources used to calculate total consumption. These approximate to 95% confidence intervals, standard across statistical analysis.

B.14 For diesel, an assessment of the effect of random variation is included using the error margins resulting from the data used to estimate illicit consumption.

B.15 The central estimate for spirits may not necessarily be half way between the upper and lower bounds as these bounds are confidence intervals, which may not be symmetric about the central estimate. As we do not have appropriate confidence intervals for the beer or tobacco tax gaps, the central estimate is calculated as the mid-point between the upper and lower estimates.

Direct taxes

Systematic biases

B.16 For direct tax estimates based on random enquiries, an adjustment is made to account for under-declarations of liabilities that are not detected. More information about our approach to non-detection multipliers can be found in HMRC’s working paper www.gov.uk/government/publications/non-detection-multipliers-for-measuring-tax-gaps. HMRC continues to undertake analysis to define suitable ranges for other systematic biases in the direct tax estimates.

B.17 Direct tax gaps that rely on management information methods measure known components separately. There are also unknown factors that are not fully identified, leading to additional unmeasured losses.

Random variation

B.18 Direct tax estimates derived from random enquiries will be subject to random sampling errors. 95% confidence intervals have been calculated for these estimates using standard statistical techniques.
Chapter C: Tax gap and compliance yield

C.1 Tax gap estimates are calculated net of compliance yield – that is, they reflect the tax gap remaining after HMRC compliance activity.

C.2 The cash expected element of compliance yield represents additional tax liabilities due which arise from checks into past non-compliance. Cash expected is tax gap closing and is part of the tax gap calculation for some but not all of the tax gap components. Because the tax gap reflects a single tax year, and some compliance cases can cover multiple tax years, the year in which cash expected is generated and recorded for compliance yield is not always the same as the year to which liabilities relate. It is possible that the amount of compliance yield HMRC secures might increase while the percentage tax gap remains the same or reduces.

C.3 To read about how the methodology for measuring tax yield in HMRC’s Annual Report and Accounts may differ from the methodology for how compliance yield is reflected in the tax gap estimates, go to www.gov.uk/government/publications/hmrc-annual-report-and-accounts-2016-to-2017.

C.4 To estimate the tax gap, some methodologies specifically use the cash collected element of compliance yield in the tax gap calculation:

<table>
<thead>
<tr>
<th>Tax gap component</th>
<th>Compliance yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self Assessment for large partnerships</td>
<td>Deducted from gross tax gap; actual compliance yield series shown in table 4.6 ‘Measuring tax gaps 2020’ edition.</td>
</tr>
<tr>
<td>PAYE (small businesses)</td>
<td>Deducted from gross tax gap; actual compliance yield series shown in table 4.8 of ‘Measuring tax gaps 2020’ edition.</td>
</tr>
<tr>
<td>PAYE (mid-sized business)</td>
<td>Deducted from gross tax gap; actual compliance yield series shown in table 4.10 of ‘Measuring tax gaps 2020’ edition. This will represent both actual compliance yield (for closed cases) and estimates of compliance yield (for tax cases which are still under enquiry).</td>
</tr>
<tr>
<td>Corporation Tax (large businesses)</td>
<td>Deducted from gross tax gap; compliance yield series shown in table 5.1 of ‘Measuring tax gaps 2020’ edition. This will represent both actual compliance yield (for closed cases) and estimates of compliance yield (for tax cases which are still under enquiry).</td>
</tr>
<tr>
<td>Corporation Tax (mid-sized businesses)</td>
<td>Deducted from gross tax gap; actual compliance yield series shown in table 5.2 of ‘Measuring tax gaps 2020’ edition. This will represent both actual compliance yield (for closed cases) and estimates</td>
</tr>
<tr>
<td>Tax gap component</td>
<td>Compliance yield</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Corporation Tax (small businesses)</td>
<td>Deducted from gross tax gap; actual compliance yield series shown in table 5.3 of ‘Measuring tax gaps 2020’ edition.</td>
</tr>
<tr>
<td>Diesel</td>
<td>Deducted from gross tax gap.</td>
</tr>
<tr>
<td>Landfill Tax</td>
<td>Deducted from gross tax gap.</td>
</tr>
</tbody>
</table>

C.5 In the following components of the tax gap we use an estimate of compliance yield as part of the calculation or do not take into account compliance yield:

<table>
<thead>
<tr>
<th>Tax gap component</th>
<th>Compliance yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avoidance (income tax, National Insurance Contributions and Capital Gains Tax)</td>
<td>Compliance yield for open cases is estimated by looking at the success of avoidance cases in a related area (large business) over time.</td>
</tr>
<tr>
<td>Hidden economy - ghosts</td>
<td>Does not currently take account of compliance yield.</td>
</tr>
<tr>
<td>Hidden economy - moonlighters</td>
<td>Based on experimental methodology which estimates the tax gap directly and does not currently take account of compliance yield.</td>
</tr>
</tbody>
</table>

C.6 In the remaining components of the tax gap we use a top-down method of calculation, looking at the difference between total theoretical liabilities and tax receipts. Although compliance yield is not explicitly included in these calculations it is reflected as part of tax receipts:

<table>
<thead>
<tr>
<th>Tax gap component</th>
<th>Compliance yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAT</td>
<td>Not explicitly used; but is reflected in receipts.</td>
</tr>
<tr>
<td>Tobacco</td>
<td>Not explicitly used; but is reflected in receipts.</td>
</tr>
<tr>
<td>Alcohol</td>
<td>Not explicitly used; but is reflected in receipts.</td>
</tr>
<tr>
<td>Stamp duties</td>
<td>Not explicitly used; but is reflected in receipts.</td>
</tr>
</tbody>
</table>

C.7 HMRC publishes a detailed breakdown of compliance revenues within our Annual Report and Accounts. This differs in coverage and timing from the compliance information presented in ‘Measuring tax gaps’. 
Chapter D: Value Added Tax

VAT gap

General methodology

D.1 The VAT tax gap is measured by estimating the total consumption of taxable goods and services to calculate the net VAT total theoretical liability (VTTL); the VAT gap is the difference between the VTTL and the VAT received. The VAT gap methodology uses a ‘top-down’ approach which involves:

- gathering data detailing the total amount of expenditure in the economy that is subject to VAT, primarily from the Office for National Statistics (ONS)
- applying the rate of VAT on the ONS expenditure data based on commodity breakdowns to derive the gross VTTL
- subtracting any legitimate refunds occurring through schemes and reliefs, to arrive at the net VTTL
- subtracting actual VAT receipts from the net VTTL
- leaving the residual element - the VAT gap, which includes, for example, error, evasion and debt

D.2 The VTTL is the amount of VAT that should be collected in theory. This means applying the rate of VAT on that expenditure where VAT should be payable, assuming that there is no fraud, avoidance, or losses due to error or non-compliance.

D.3 The VTTL includes irrecoverable VAT, which is the VAT paid on ‘finally taxed expenditure’ which cannot be reclaimed, for example by those not registered for VAT.

D.4 The expenditure data series used in the calculation are mainly constituents of National Accounts macroeconomic aggregates. All National Accounts data used to construct VTTL estimates is consistent with the ONS Blue Book 2019.

D.5 For more information about the consumer expenditure data sources go to www.ons.gov.uk/economy/grossdomesticproductgdp.

Calculation of gross VTTL

D.6 The gross VTTL is calculated by multiplying the total amount of expenditure in the economy (also known as VAT-able expenditure) by the appropriate VAT rates.

D.7 For each of the expenditure sectors, the total expenditure is split according to the different VAT treatments; zero rated, standard rated, reduced rated, and exempt. For the purposes of calculating the gross VTTL, only the standard and reduced rated expenditure are used.

D.8 The total VAT-able expenditure for each sector is combined together to represent an overall annual figure for the economy.

D.9 In order to derive the amount of VAT within the VAT-able expenditure, it is necessary to multiply the expenditure by the VAT fraction. The annual gross VTTL is thus calculated by multiplying the annual expenditure figure for the economy by the respective VAT fraction.
D.10 A number of streams of expenditure contribute to the tax base, with most VAT deriving from consumers’ expenditure (that is, household consumption). The main expenditure categories that comprehensively cover VAT liabilities are:

- household consumption
- non-profit institutions serving households
- government capital and current expenditure
- VAT exempt sector capital and current expenditure
- housing capital expenditure

**Input tax adjustments**

D.11 Net VAT liability is the difference between VAT due on taxable supplies made by registered traders (‘output tax’), and VAT recoverable by traders on supplies made to them (‘input tax’).

D.12 VAT liability for the relevant categories can be estimated directly from ONS National Accounts data, with one exception – the VAT exempt sector. Businesses making outputs that are exempt from VAT are generally not permitted to reclaim all the VAT on inputs associated with their exempt outputs. In order to make an adjustment for this irrecoverable input tax, a separate HMRC survey is used to ascertain the proportion of purchases on which VAT cannot be reclaimed.

D.13 A further adjustment is made for expenditure by businesses which are legitimately not registered for VAT and, as such, cannot recover their input tax. This adjustment uses a combination of data from the Department for Business, Energy and Industrial Strategy (BEIS) and HMRC information on the distribution of business turnover below the VAT threshold to estimate relevant expenditure.

D.14 Finally, HMRC data and third-party data sources are used in conjunction with National Accounts data to inform estimates of business expenditure on cars and entertainment, on which VAT is due.

D.15 Because the calculation of irrecoverable input tax is complex, the level of uncertainty around input tax adjustments is larger than for the other elements.

**Deductions**

D.16 The sum of the VAT liability arising from each of the expenditure categories listed in paragraph D.10 gives an estimate of the gross VTTL in each year. However, there are a number of legitimate reasons why part of this theoretical VAT is not actually collected. These can be grouped into three broad categories:

- VAT refunds
- expenditure of traders legitimately not registered for VAT
- other deductions

D.17 VAT refunds are made primarily to government departments, NHS Trusts and regional health authorities for specified contracted out services acquired for non-business purposes. A number of other categories of expenditure cannot be separately identified in the overall VTTL calculation, for which VAT can be refunded. The value of these refunds is taken directly from audited HMRC accounts data.

D.18 Traders who trade below the VAT threshold can legitimately exclude VAT on their sales. Expenditure on the output of these businesses will have been picked up in the total theoretical
liability. To adjust for this, an estimate of relevant expenditure is made using a combination of BEIS data and HMRC information on the distribution of business turnover below the VAT threshold.

D.19 Other deductions include, for example, legitimate schemes (such as the Retail Export Scheme) and reliefs.

Net VAT receipts

D.20 Figures for actual receipts of VAT are taken from HMRC’s published National Statistics tax receipts figures. The receipts are adjusted to reflect timing effects within each tax year, before being used in the model. A summary of HMRC’s tax receipts can be found here [www.gov.uk/government/statistics/hmrc-tax-and-nics-receipts-for-the-uk](http://www.gov.uk/government/statistics/hmrc-tax-and-nics-receipts-for-the-uk).

VAT gap

D.21 Finally, subtracting the Net VAT Receipts from the Net VTTL derives the VAT gap. The percentage gap is further calculated by dividing the VAT gap by the Net VTTL. Receipts for the tax year (April to March) are compared with the total theoretical liability for the calendar year, assuming an average three-month lag between an economic activity and the payment of the corresponding VAT to HMRC. Calculations for VTTL and Net VTTL assume a three-month lag between expenditure and actual VAT receipts. Hence, calendar year expenditure data equates to tax year receipts.

D.22 The detailed calculations used to construct the estimated VTTL are continuously reviewed to identify improvements to the methodology. Also, the National Accounts data used to construct the VTTL is subject to updates and revision by the ONS throughout the year. This is part of the routine revisions to the ONS National Accounts data as final data become available.

D.23 The flowchart below illustrates the series of model operations described above using symbols to represent each step of the process and contains a short description of the process step:
Spirits and beer (upper bound) estimate

Overview

E.1 The estimates of the illicit market for spirits and the beer upper bound are produced using a top-down methodology. That is, the estimate is produced by first estimating total consumption, and then subtracting legitimate consumption, with the residual being the illicit market:

\[ \text{Illicit market} = \text{Total consumption} - \text{Legitimate consumption} \]

E.2 The above equation provides an estimate of the volume of goods supplied through the illicit market. This is then turned into an estimate of the proportion of the total market that is supplied through the illicit market, termed the illicit market share:

\[ \frac{\text{Illicit market}}{\text{Total consumption}} \times 100 \]

E.3 Revenue losses associated with the illicit market are then estimated by combining the illicit market share information with price data, excise duty and VAT rate information.

E.4 Although the spirits and the beer upper bound estimates are calculated using the same underlying methodology there are differences, the two main ones being

- the spirits tax gap estimate uses one methodology and is produced with confidence intervals, whilst beer has two methodologies: an upper and a lower bound estimate which are averaged to produce an implied midpoint estimate
- the spirits and beer estimates use different methods to calculate the uplift factors

E.5 Details of the methodology, including differences, for the estimation of the spirits and beer (upper bound) tax gap are provided in the next sections, followed by the lower bound beer tax gap.

Estimating total consumption

E.6 The consumption of spirits or beer bought in the United Kingdom (UK) is estimated using the Family Spending Surveys (FSS) from the Office for National Statistics (ONS). This includes the Living Costs and Food Survey (LCF). LCF estimates are weighted by the ONS to adjust for survey non-response.

E.7 Since the FSS only covers purchases within the UK, cross-border shopping is added to the consumption of spirits/beer bought in the UK to give total consumption.

Total consumption of UK purchases

E.8 The consumption of UK purchased goods in any given year is calculated using the following:

- estimates of household on-licence (consumed at the point of sale, for example, in a pub or restaurant) and off-licence (consumed off the premises, for example from a supermarket) expenditure on spirits/beer from the FSS
• the average number of people in a household estimated from the FSS
• data on average alcohol prices provided by the ONS
• estimates of the UK adult population (ages 18 or over) from the ONS
• uplift factors calculated independently for on-licence and off-licence sectors

E.9 Total UK consumption is defined using this formula:

\[
\text{Total UK consumption} = \frac{\text{Average household consumption}}{\text{Average number of adults per household}} \times \text{UK adult population} \times \text{Uplift factor}
\]

Family Spending Surveys

E.10 The average weekly expenditure on spirits and beer for an average UK household is estimated using the FSS. Households participating in the surveys are asked to record their expenditure on alcohol under the relevant specific category of drink (that is wine, spirits, beer, etc.). There is an additional category for recording drinks purchased as part of a ‘round’ of drinks, which will be referred to as ‘other drinks’.

E.11 Some of the ‘other drinks’ consumed will be spirits or beer. The calculation for consumption therefore includes a proportion of ‘other drinks’ consumption.

E.12 The average weekly expenditure per household is converted to the volume consumed by that household using the average price of spirits/beer. This is then scaled up to an annual figure.

E.13 The average consumption of spirits/beer per household is then converted to the average per person, by dividing by the average number of adults in a household. This is scaled up to the UK adult population.

E.14 Most under-age drinking is taken into account in the alcohol models. We assume that adults buy most of the alcohol consumed by minors. This under-age alcohol expenditure is therefore included in the adults’ alcohol consumption and is measured by the survey.

E.15 Due to the relatively small sample size in the FSS, the average weekly expenditure for spirits or beer is heavily influenced by extreme expenditure values in the data. Outliers in the data have been capped at the 99th percentile.

Cross-border shopping and duty free

E.16 Duty free is included in the cross-border shopping calculation. Estimates of consumption of goods purchased as cross-border shopping are based on figures produced from the International Passenger Survey (IPS). This provides estimates of the volume of spirits an average adult traveller brings into the country, separately for air and sea passengers. The IPS figures are weighted by the ONS, scaling up the survey data to represent the total cross-border shopping entering the UK.

E.17 An estimate of the volume of duty-free spirits/beer brought into the country is calculated in the same way, using passengers coming from outside the European Union (EU).

E.18 This estimate, however, does not cover sales made on-board ferries, so commercially provided data about deliveries of spirits/beer to ferries are used to supplement the cross-border shopping estimate, and provide a complete figure.

E.19 Cross-border shopping is estimated as:
Estimating legitimate consumption

E.20 Estimates of legitimate consumption have two elements; UK duty paid consumption and cross-border shopping:

\[
\text{Legitimate consumption} = \text{UK duty paid consumption} + \text{Cross-border shopping}
\]

E.21 Estimates of UK duty paid consumption are taken directly from returns to HMRC of the volumes of spirits/beer on which duty has been paid. For spirits the volumes of ready-to-drink products have been removed from spirits clearances in order to obtain figures for spirits only.

E.22 Cross-border shopping is calculated in the same way as for total consumption:

\[
\text{Cross-border shopping} = \text{Goods bought overseas} + \text{Goods bought on-board ferries} + \text{Duty free}
\]

Estimating the illicit market

E.23 Combining the two formulae for total consumption and legitimate consumption gives the formula for the illicit market:

\[
\text{Illicit market} = \text{Total consumption} - \text{Legitimate consumption}
\]

Conversion to monetary losses

E.24 Revenue losses associated with the illicit market are then estimated by combining the illicit market share information with price data and duty and VAT rate information, using this formula:

\[
\text{Financial losses} = \left( \text{Illicit volume} \times \frac{\text{Spirits/beer duty}}{\text{Average price}} \times \text{VAT fraction} \right)
\]

E.25 Data on average spirits/beer prices is derived from data provided by the ONS. The prices used in the model are weighted across on and off licence and for different types of spirits/beer.

E.26 The VAT fraction is the portion of the retail price that is VAT – for example, a 20% VAT rate is equivalent to a one-sixth VAT fraction. VAT fractions are calculated annually to capture changes in
the VAT rate. This method assumes that VAT is also lost on all purchases. As, in some cases, the final illicit product is sold in legitimate outlets this may not always be the case, and this will be an overestimate of losses.

E.27 For the spirits calculation, spirits duty is converted into bulk duty liabilities based on the assumption that spirit's strength is constant at 38%.

Summary of methodology

E.28 The flowchart below illustrates the series of model operations described above using symbols to represent each step of the process and contains a short description of the process step to calculation the illicit market estimate using spirits as an example:

\[
\text{Illicit market} = \text{Total consumption} - \text{Legitimate consumption}
\]

\[
\text{UK duty paid consumption} + \text{Cross-border shopping}
\]

\[
(\text{Average adult spirits consumption} \times \frac{\text{UK adult population}}{\text{Uplift factor}}) + (\text{Proportion of average adult 'other' drinks consumption} \times \frac{\text{UK adult population}}{\text{Uplift factor}})
\]

Cross-border shopping

Spirits uplift factor

E.29 The LCF Survey data for alcohol are subject to under-reporting, they may under-represent certain sub-populations with a high average alcohol consumption, and do not cover the full extent of the alcohol market so an uplift factor is necessary to correct for this bias. This uplift factor is calculated by taking estimates of consumption from the LCF Survey in the base year and comparing these with independent estimates of total consumption.

E.30 To do this we take a year in which there is believed to be little or no illicit market and use HMRC clearance data as a true indication of total consumption. In order to reduce sampling error, the uplift factor is derived by taking the average of three year's data: 1990 to 1991, 1991 to 1992 and 1992 to 1993.

E.31 Separate uplift factors are calculated for on-license and off-licence markets, and the formula is defined as:

\[
\text{Base year uplift factor} = \frac{\text{Legitimate consumption in 1990-91 to 1992-93}}{\text{Estimated total consumption in 1990-91 to 1992-93}}
\]

E.32 The uplift factors for on-licence and off-licence are 3.5 and 2.0 respectively.

Beer Uplift factor

E.33 The basis for this uplift factor is the same as for spirits, an average of three base years is used where there is assumed to be no illicit market. However due to the variation in price between draught and packaged beer, a different uplift factor to spirits is required.
To calculate uplift factors for draught and packaged beer, LCF Survey data is split between on-licence and off-licence markets and then into draught and packaged beer. This uses market shares estimated from ONS and British Beer and Pub Association (BBPA) data.

The base year uplift factors are defined as:

\[
\text{Base year uplift factor} = \frac{\text{Legitimate consumption in 1990-91 to 1992-93}}{\text{Estimated total consumption in 1990-91 to 1992-93}}
\]

An additional uplift for packaged beer is calculated, which varies year-on-year. This assumes that there is no or negligible illicit market in draught beer, whereby consumption is equal to clearances in every year. The draught beer uplift and base year uplifts are combined to compute the packaged beer uplift, providing a 3.1 uplift for 2018 to 2019, using this formula:

\[
\text{Packaged uplift} = \text{Draft uplift} \times \frac{\text{Base year packaged uplift}}{\text{Base year draft uplift}}
\]

Removing spirit-based ready-to-drinks

Spirit-based ready-to-drinks (RTDs) are packaged beverages that are sold in a prepared form, ready for consumption, such as alcopops.

The LCF survey expenditure data for spirits includes expenditure on RTDs.

RTDs are currently included in the ‘other excise duties’ estimates, so are removed from the spirits tax gap to avoid double counting. To remove RTDs, we estimate the proportion of total expenditure attributable to ready-to-drinks using data on expenditure from the ONS, and total pure alcohol clearances on spirits and RTDs from HMRC clearances.

Upper and lower confidence intervals in spirits estimate

The variation in the FSS is used to construct 95% confidence intervals around the central estimate. They indicate the potential size of chance fluctuations in the estimate due to sampling error. They do not take into account systematic error from the model assumptions in the central estimate.

Beer lower estimate

Overview

The beer tax gap lower estimate is produced using a bottom-up methodology. This means estimates of the illicit market are made directly, by estimating the fraud components that make up the illicit market. The following types of illicit beer are included in the lower estimate:

- diversion of UK-produced beer
- drawback fraud

Some of this illicit beer is recovered through HMRC compliance activity, so this is subtracted to give the net tax gap. The tax gap estimate is defined by:

\[
\text{Beer illicit market lower estimate} = \text{Diversion of UK produced beer} + \text{Drawback fraud} - \text{Seizures of illicit beer}
\]
E.43 A number of beer frauds are not included in this methodology as we are currently unable to estimate them. This is one of the reasons it is a lower bounding estimate. These include:

- smuggled beer
- diversion of foreign produced beer
- counterfeit beer
- any other fraud we do not know about

**Diversion of UK-produced beer**

E.44 Diversion fraud occurs when beer is moved in duty suspense to the EU and is subsequently diverted back into the UK under the cover of false documentation. The taxes are not declared on the beer and the illicit product enters the UK market.

E.45 We estimate that diversion fraud is equal to the amount of beer moved in duty suspense from the UK to certain EU member states, minus legitimate demand for UK branded beer in those countries. That is, we assume that any UK beer which is not feeding demand abroad will be diverted back to the UK illicit market:

\[
\text{Diversion of UK produced beer} = \text{Duty suspended beer moved to selected EU countries} - \text{Legitimate demand in selected EU countries}
\]

E.46 The total amount of beer moved in duty suspense from the UK to the EU includes dispatches from both excise warehouses and brewers. Dispatches from excise warehouses are taken directly from Excise Warehouse Returns (W1 form). Dispatches from brewers are estimated using data from Beer Duty Returns (EX46 form):

\[
\text{Total beer dispatches} = \text{Dispatches from warehouses} - \text{Dispatches from brewers}
\]

E.47 Brewers return data is used for dispatches (movements to EU countries) and exports (movements to non-EU countries) and it cannot be disaggregated. So, to estimate dispatches from brewers, we subtract an estimate of exports from brewers.

E.48 Exports from brewers are estimated as total exports, from Customs Handling of Import and Export Freight (CHIEF), minus exports from Excise Warehouse Returns (W1 form). Note that Excise Warehouse Returns have been forecast for the tax year 2018 to 2019.

E.49 To preserve the lower bounding nature of this estimate, we only include dispatches to certain EU countries. These countries have been selected based on a number of factors, including: proximity to the UK; the differential in price; operational indications of risk and patterns of supply.
The estimate of beer dispatches, described in paragraph E.47 and E.49, cannot be broken down to the recipient country. Therefore, we use an alternative data source, UK trade data, which does include a breakdown by country. The proportion of beer dispatched to the selected EU countries is taken from UK trade data and applied to the estimated total dispatches to produce an estimate for dispatches to these selected EU countries.

UK trade data is not used to directly estimate dispatches to these countries as it does not include certain types of movements. More detail is provided on this in paragraph E.65.

The flowchart below illustrates the series of model operations described above using symbols to represent each step of the process and contains a short description of the process step:

\[
\text{Duty suspended beer moved to selected EU Countries} = \text{Total dispatches to EU countries} \times \text{Percentages of dispatches going to selected EU countries}
\]

\[
\text{Dispatches from warehouses + Dispatches from brewers}
\]

\[
\text{Dispatches and exports from brewers} - \text{Exports from brewers}
\]

\[
\text{Total exports} - \text{Exports from warehouses}
\]

**Drawback fraud**

Drawback fraud occurs when goods are moved to the EU and the duty is reclaimed via drawback. Duty is then paid at the lower rate in the destination country and the goods are illicitly returned to the UK.

To estimate drawback fraud, we estimate the volume of beer corresponding to certain drawback claims, then subtract the legitimate demand for beer in the selected destination countries:

\[
\text{Drawback fraud} = \frac{\text{Selected drawback claims converted to volume of beer}}{\text{selected EU countries}} - \frac{\text{Legitimate demand in selected EU countries}}{\text{}}
\]

To preserve the lower bounding nature of this estimate, we only include drawback if it is claimed for dispatch by a business not part of HMRC Large Business (LB). The value of these drawback claims are converted to volumes of beer by dividing by the average duty rate for beer.

The volume is then adjusted using the proportion of dispatches going to the selected EU countries. This gives an estimate of the amount of beer going to the selected countries with drawback claimed by small and medium sized enterprises:
Legitimate demand in selected EU countries

E.57 Some of the beer moved to the selected EU countries will be supplying legitimate demand within those countries, rather than being diverted to the UK illicit market. We make one overall estimate of legitimate demand in the selected EU countries and subtract it from the sum of selected beer dispatches and selected beer for drawback.

E.58 We have purposely overestimated legitimate demand by only accounting for the riskiest countries, which produces an underestimate of the illicit market, in order to maintain the lower bounding nature of the tax gap estimate.

E.59 The estimate of legitimate demand in other countries includes:

- cross-border shopping bought by UK residents
- legitimate consumption abroad, which may include: consumption by UK expatriates, consumption by UK residents while abroad, consumption by foreign nationals, beer in transit to other countries

\[
\text{Legitimate demand in selected EU countries} = \text{Cross-border shopping by UK residents} + \text{Legitimate consumption abroad}
\]

E.60 Cross-border shopping is estimated using data from the IPS. More detail is provided in paragraph E.16. Only passengers from the selected EU countries are included.

Legitimate consumption of UK produced beer abroad

E.61 We could not find reliable data on legitimate consumption of UK produced beer abroad. So, we estimate it based on the assumption that in a certain year, when the illicit market upper estimate was low, there was negligible illicit activity meaning all dispatches to the selected EU countries were consumed legitimately. This is likely to provide an overestimate of legitimate consumption abroad, as there would likely be some level of fraud in these years. This supports the methodology being a lower estimate of the tax gap.

E.62 For stability, an average of two years is used: 2000 to 2001 and 2001 to 2002. We refer to these two years as the ‘base year’.

E.63 Brewers return data is not available for years prior to 2007. Consequently, we use an alternative data source, UK trade data, to estimate dispatches in the base year.

E.64 In the base year we assume that all dispatches supply either cross-border shopping by UK residents or legitimate consumption abroad. We subtract an estimate of cross-border shopping in the base year from dispatches in the base year; the remainder is assumed to be legitimate consumption abroad:
E.65 We believe that UK trade data may underestimate beer dispatches in the base year as it does not record certain types of beer movement. These include:

- goods in transit
- deliveries to embassies
- deliveries to Navy, Army and Air Force Institutes (NAAFI)

E.66 Additionally, as the threshold for recording goods on UK trade data is relatively high in beer terms, beer may have a higher proportion of small traders than other commodities. This may mean the standard adjustment applied to UK trade data to account for small traders may be too low for beer.

E.67 To account for these concerns, we uplift the UK trade data. There is very little evidence to indicate the actual scale of uplift required. Comparison with our calculated dispatches in later years led us to apply a factor of two. Again, the high level of this adjustment may result in this being an overestimate, but this is in keeping with the lower bounding methodology for the tax gap.

**Illicit market lower estimate**

E.68 The flowchart below illustrates the series of model operations described above using symbols to represent each step of the process and contains a short description of the process step to calculate the illicit market estimate:
**Implied mid-point estimate**

E.69 The implied mid-point estimate is calculated as the average of the upper and lower estimates. It is only intended as an indicator of long-term trend – the true tax gap could lie anywhere within the bounds.

E.70 The bounds do not take account of any systematic tendency to over- or under-estimate the size of the tax gap that might arise from the modelling assumptions.

**Wine central estimate**

E.71 We have not estimated the illicit market share for wine due to the unavailability of a key underlying commercial data source previously used to estimate the wine tax gap.

E.72 We previously produced an illustrative estimate of the wine tax gap, from 2013 to 2014, to 2016 to 2017 inclusive, by taking the average illicit market share from 2011 to 2012, and 2012 to 2013. These illustrative estimates were static and did not reflect the continuing changes within the alcohol market and compliance. We have therefore stopped publishing this estimate. We have now included wine within our illustrative tax gap estimate for ‘Other indirect taxes’, see Chapter J Other taxes.
Overview

F.1 The estimate of the illicit market for tobacco is produced using a top-down methodology. That is, first we estimate total consumption, and then we subtract legitimate consumption. The residual is estimated to be the illicit market.

\[
\text{Illicit market} = \text{Total consumption} - \text{Legitimate consumption}
\]

F.2 The above equation provides an estimate of the volume of goods supplied through the illicit market. This is then turned into an estimate of the proportion of the total market that is supplied through the illicit market.

\[
\text{Illicit market share} = \frac{\text{Illicit market}}{\text{Total consumption}} \times 100
\]

F.3 Revenue losses associated with the illicit market are then estimated by combining the illicit market share information with price data, excise duty, and VAT rate information.

Methodology

F.4 The estimates of the illicit market for cigarettes and hand-rolling tobacco (HRT) are produced using a top-down methodology as described in paragraphs F.1 to F.3. These estimates combined provide the tobacco tax gap.

F.5 Details of the estimation of total consumption and of legitimate consumption are provided in the subsequent sections.

F.6 Due to changes to the Office of National Statistics’ Opinions Survey, which is used to estimate total consumption, the percentage tobacco gap for this year has been projected from the 2017 to 2018 tax year. This means that the percentage tax gap has been kept the same, and total consumption has been scaled according to total clearances.

Total consumption

F.7 The total consumption in any given year is calculated using the following:

- estimates of prevalence (proportion of the population that smokes cigarettes) from the General Lifestyle Survey (GLF), the Opinions and Lifestyle Survey (OPN) and Health Survey for England (HSE)
- estimates of cigarette consumption per smoker from GLF, OPN and HSE
- estimates of the adult population (ages 16 or over) from the Office for National Statistics (ONS)
- an uplift factor covering under-reporting

F.8 The estimate of total UK consumption of cigarettes and HRT for each year is a product of the estimates of cigarette and HRT smoking prevalence and consumption per smoker for declared and undeclared smokers.

F.9 In general, most smokers admit that they smoke but there are some smokers who, for whatever reason, do not admit that they smoke. We therefore obtain the prevalence and consumption per smoker of these declared smokers from the OPN since 2012 and obtain the undeclared smokers in the non-smoking population from the HSE.
Uplift factor

F.10 The survey data for tobacco consumption is subject to under-reporting due to the self-reporting nature of the surveys. An uplift factor is necessary to correct for this bias. This uplift factor is calculated by taking estimates of total consumption from the GLF in a base year, and comparing with consumption based on actual clearances to HMRC and an estimate of legitimately purchased cigarettes from abroad.

F.11 The uplift factors for the 2018 to 2019 cigarettes and hand-rolling tobacco estimates are 1.5 and 1.1 respectively. These were calculated from the base year using the following formula:

\[
\text{Uplift factor} = \frac{\text{Legitimate consumption (from HMRC clearances and estimate of duty free/cross border shopping)}}{\text{Total consumption (based on self-reported consumption from GLF survey)}}
\]

Upper and lower bounds for total consumption

F.12 The uncertainties in the survey data used to create these estimates mean that it is not possible, with sufficient accuracy, to produce a single point estimate of total consumption. However, due to the methodology we use, it is difficult to produce confidence intervals so instead, we carry out an exercise based on the survey data to produce an upper bound and lower bound for total consumption. This allows us to produce a range for total consumption that takes account of the uncertainty in the underlying data.

F.13 The one difference between the upper and lower bound calculations is the treatment of dual smokers. In the upper bound calculation, the majority of the dual smokers are considered to be cigarette smokers. In the lower bound estimate, we assume that the majority of them smoke hand-rolling tobacco. This is explained further in the following tables and sections.

Table F.1 Cigarettes upper and HRT lower bound assumptions

<table>
<thead>
<tr>
<th>OPN Survey Options</th>
<th>Allocation of total tobacco consumption for estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cigarette upper bound assumption</td>
</tr>
<tr>
<td>Cigarettes only</td>
<td>100%</td>
</tr>
<tr>
<td>Dual smokers: cigarettes and hand-rolling tobacco, but mainly cigarettes</td>
<td>99%</td>
</tr>
<tr>
<td>Dual smokers: cigarettes and hand-rolling tobacco, but mainly hand-rolling tobacco</td>
<td>49%</td>
</tr>
<tr>
<td>Hand-rolling tobacco only</td>
<td>0%</td>
</tr>
</tbody>
</table>
Table F.2 Cigarettes lower and HRT upper bound assumptions

<table>
<thead>
<tr>
<th>OPN Survey Options</th>
<th>Cigarette lower bound assumption</th>
<th>HRT upper bound assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cigarettes only</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>Dual smokers: cigarettes and hand-rolling tobacco, but mainly cigarettes</td>
<td>51%</td>
<td>49%</td>
</tr>
<tr>
<td>Dual smokers: cigarettes and hand-rolling tobacco, but mainly hand-rolling tobacco</td>
<td>1%</td>
<td>99%</td>
</tr>
<tr>
<td>Hand-rolling tobacco only</td>
<td>0%</td>
<td>100%</td>
</tr>
</tbody>
</table>

F.14 The upper bound of total cigarette or HRT consumption is calculated firstly by estimating consumption levels from smokers who only smoked cigarettes or HRT. This is added together with a maximum consumption of cigarettes or HRT that could be smoked by dual smokers.

F.15 The lower bound of total cigarette or HRT consumption is calculated firstly by estimating consumption levels from smokers who only smoked cigarettes or HRT. This is added together with a minimum consumption of cigarettes or HRT that could be smoked by dual smokers.

F.16 Prior to 2012, the GLF was used as the base estimate for cigarette consumption. This was supplemented with OPN data on dual smokers where this was added/subtracted to obtain the upper and lower bounds.

Legitimate consumption

F.17 Estimates of legitimate consumption include:
- UK duty paid consumption
- Cross-border and duty-free shopping

UK duty paid consumption

F.18 Estimates of UK duty paid consumption are taken directly from tax returns to HMRC (clearance data) on the volumes of cigarettes and HRT on which duty has been paid, along with the actual amounts of money.

Cross-border shopping and duty free

F.19 Estimates of consumption of goods purchased as cross-border shopping are based on data from the International Passenger Survey (IPS). This provides estimates of the number of cigarettes and/or HRT that an average adult traveller brings into the country, separately for air and sea passengers. The IPS figures are weighted by the ONS, scaling up the survey data to represent the total cross-border shopping entering the UK.

F.20 This estimate, however, does not cover sales made on-board ferries, so commercially provided data about deliveries of cigarettes to ferries is used to supplement the cross-border shopping estimate.

F.21 Duty free cigarettes/hand-rolling tobacco brought into the UK are also estimated from the IPS, using passengers coming back from outside the EU.
F.22 Legitimate consumption can be summarised as:

\[
\text{Legitimate consumption} = \text{UK duty paid consumption} + \text{Cross-border shopping} + \text{Duty free}
\]

**Conversion to monetary losses**

F.23 Revenue losses associated with the illicit market are then estimated by combining the illicit market share information with price data, duty, and VAT rate information.

F.24 All calculations to this point have been made on volumes of cigarettes or HRT. Volumes are converted to estimates of revenue using:

\[
\text{Losses} = (\text{Specific duty} + (\text{Ad valorem duty} + \text{VAT fraction}) \times \text{Average price}) \times \text{Illicit volume}
\]

F.25 The average price is taken as the weighted average price (WAP) of all cigarettes or HRT that were UK duty paid. The WAP is calculated by weighting the retail price of each product by the share of clearances in the cigarette or HRT market.

F.26 The VAT fraction is the proportion of the retail price that is VAT – for example, a 20% VAT rate is equivalent to one-sixth VAT fraction. VAT fractions are calculated annually to capture changes in the VAT rate. This method assumes that VAT is also lost on all purchases. In some cases, the final illicit product is sold in legitimate outlets this may not always be the case, and this will be an overestimate of losses.

**Summary of cigarette methodology**

F.27 The flowchart below illustrates the series of model operations described above using symbols to represent each step of the process and contains a short description of the process step to calculate the illicit market for the cigarettes tax gap:
Summary of HRT methodology

F.28 The flowchart below illustrates the series of model operations described above using symbols to represent each step of the process and contains a short description of the process step to calculate the illicit market for the HRT tax gap:

\[
\text{Illicit market} = \text{Declared consumption} + \text{Undeclared consumption} - \text{UK duty paid consumption} - \text{Cross border shopping} - \text{Duty free}
\]

\[
\text{HMRC Clearances data} \quad \text{From IPS}
\]

\[
\text{Average amount per traveller} \times \text{Number of travellers} + \text{On-board ferry sales}
\]

\[
\text{Non smoker adult population} \times \text{Uplift factor} \times \text{Under-declared smokers prevalence} \times \text{Consumption per under-declared smoker}
\]

\[
\text{Adult population} \times \text{Uplift factor} \times \left( \text{Declared consumption by 'cigarette only' respondents} + \text{Declared consumption from dual smokers assuming most smoke [tobacco product]} \right)
\]

\[
\text{[tobacco product]} = \text{Cigarettes [upper bound]} \ OR \ HRT [lower bound]
\]
Illicit market = Declared consumption + Undeclared consumption - UK duty paid consumption - Cross border shopping - Duty free

\[ \text{Illicit market} = \text{Declared consumption} + \text{Undeclared consumption} - \text{UK duty paid consumption} - \text{Cross border shopping} - \text{Duty free} \]

\[ \text{HMRC Clearances data} \]

\[ \text{From IPS} \]

\[ \text{Average amount per traveller} \times \text{Number of travellers} + \text{On-board ferry sales} \]

\[ \text{Non smoker adult population} \times \text{Uplift factor} \times \text{Under-declared smokers prevalence} \times \text{Consumption per under-declared smoker} \]

\[ \text{Adult population} \times \text{Uplift factor} \times (\text{Declared consumption by 'HRT only' respondents} + \text{Declared consumption from dual smokers assuming most smoke [tobacco product]}) \]

[tobacco product] = HRT [upper bound] OR Cigarettes [lower bound]
Methodology

G.1 A bottom-up methodology is used to estimate the diesel tax gap from tax year 2016 to 2017 onwards based on a random enquiry programme. The Great Britain (GB) and Northern Ireland (NI) diesel tax gaps are calculated separately but the methodologies are identical. Figures from 2016 to 2017 onwards are not directly comparable to previous years, as they are based on a top-down methodology.

G.2 Summary of methodology:
- legitimate consumption is based on the returns that HMRC receives from the volumes of diesel on which duties have been paid (HMRC clearances)
- illicit consumption is estimated using the proportion of vehicles found to be misusing rebated fuel in a random sample survey conducted by HMRC in 2017
- revenue losses (gross tax gap) associated with illicit consumption are estimated using average retail prices, duty rates and VAT rates
- the net tax gap is then calculated as the gross tax gap minus compliance yield

Estimating total consumption

G.3 In April to June 2017, HMRC conducted a random survey where vehicles were stopped at the roadside and tested for illicit diesel. A stratified sample of 1,900 vehicles across the UK (1,500 in GB and 400 in NI) was used. The sample was stratified by vehicle type and region to ensure the results were representative.

G.4 The proportion of vehicles found to be misusing rebated fuel (strike rate) in this survey is used as an estimate of the proportion of vehicles misusing rebated fuel in the UK. The strike rate is then used alongside legitimate consumption to give estimates for total and illicit consumption. A separate strike rate is calculated for GB and NI.

G.5 The flowchart below illustrates the series of model operations using symbols to represent each step of the process and contains a short description of the process step to calculate total diesel consumption:

\[
\text{Total consumption} = \text{Legitimate consumption} + \text{Illicit consumption}
\]

\[
\text{HMRC clearances} \times \frac{\text{Strike rate}}{1 - \text{strike rate}}
\]

\[
\text{Strike rate} \times \frac{\text{Number of vehicles found to be misusing rebated fuel}}{\text{Number of vehicles tested}}
\]
Conversion to monetary losses

G.6 The diesel tax gap is driven by the misuse of rebated fuel. Rebated fuel is subject to a lower duty rate and has a lower retail price including VAT. Revenue loss occurs where this fuel is misused, and so should have been subject to a higher rate of fuel duty and additional VAT.

G.7 In order to estimate the revenue losses associated with the misuse of rebated fuel, the duty and VAT paid needs to be taken into account. Therefore, the difference between rebated and un-rebated duty rates has been used to estimate the duty loss associated with the illicit market.

G.8 Similarly, the difference in average retail prices for rebated fuel and un-rebated diesel has been used to estimate the VAT loss associated with the illicit market. Published data from the Department for Business, Energy and Industrial Strategy (BEIS) has been used to calculate average retail prices.

Confidence intervals

G.9 The upper and lower estimates correspond to confidence intervals that indicate the range where the true value of the illicit market may lie and arises due to random sampling error in calculating the strike rate.

Exclusions

G.10 Smuggling of diesel is excluded on the basis that it is believed to be a minor issue, the scale of which isn’t currently quantifiable. Cross-border shopping is excluded due to a reduced-price difference between the Republic of Ireland and NI, meaning there is limited motivation for cross-border shopping activities. Revenue losses are assumed to be related to the misuse of gas oil (red diesel) only. The misuse of other fuels (for example, kerosene) have been excluded on the basis that this is believed to be a minor issue, the scale of which isn’t currently quantifiable.
Chapter H: Estimates using random enquiry programmes

H.1 This chapter covers all the approaches taken to produce income tax (IT), National Insurance Contributions (NICs) and Capital Gain Tax (CGT) tax gaps as well as the Corporation Tax (CT) and Employer Compliance (EC) gap from small business. The EC gap for large business employers is based on trends in the small business gap, the details of this are described in paragraph H22.

Random enquiry programme estimates

H.2 There are three direct tax random enquiry programmes which are used to produce tax gap estimates. They cover:

- Self Assessment (SA) individuals and small partnerships
- small business employers
- CT for small businesses

H.3 Random enquiry programmes allow HMRC to estimate the extent of under-declaration of liabilities arising from the submission of incorrect returns. Each return selected is subject to a full enquiry involving a complete examination of records. Under certain circumstances, a full enquiry may not take place if the return can be verified through third party information.

Populations and sampling

H.4 The sizes of the samples for the three programmes are shown in Table H.1 below.

Table H.1: Sample sizes for the Self Assessment, employer compliance and Corporation Tax random enquiry programmes

<table>
<thead>
<tr>
<th></th>
<th>Self Assessment</th>
<th>Employer Compliance</th>
<th>Corporation Tax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tax return</td>
<td>Sample size</td>
<td>Tax return year</td>
<td>Sample size</td>
</tr>
<tr>
<td>year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-06</td>
<td>5,234</td>
<td>2005-06</td>
<td>1,303</td>
</tr>
<tr>
<td>2006-07</td>
<td>2,925</td>
<td>2006-07</td>
<td>1,195</td>
</tr>
<tr>
<td>2007-08</td>
<td>2,864</td>
<td>2007-08</td>
<td>1,088</td>
</tr>
<tr>
<td>2008-09</td>
<td>2,708</td>
<td>2008-09</td>
<td>1,197</td>
</tr>
<tr>
<td>2009-10</td>
<td>2,116</td>
<td>2009-10</td>
<td>1,263</td>
</tr>
<tr>
<td>2010-11</td>
<td>2,028</td>
<td>2010-11</td>
<td>613</td>
</tr>
<tr>
<td>2011-12</td>
<td>2,228</td>
<td>2011-12</td>
<td>671</td>
</tr>
<tr>
<td>2012-13</td>
<td>2,170</td>
<td>2012-13</td>
<td>693</td>
</tr>
<tr>
<td>2013-14</td>
<td>2,032</td>
<td>2013-14</td>
<td>760</td>
</tr>
<tr>
<td>2015-16</td>
<td>2,032</td>
<td>2015-16</td>
<td>715</td>
</tr>
<tr>
<td>2016-17</td>
<td>2,248</td>
<td>2016-17</td>
<td>619</td>
</tr>
<tr>
<td>2017-18</td>
<td>2017-18</td>
<td>675</td>
<td>2017-18</td>
</tr>
</tbody>
</table>

Notes for Table H.1:
1. Since the tax year 2015 to 2016 the Employer Compliance sample size given is for small business only
2. Since the tax year 2016 to 2017 the Corporation Tax sample size given is for small business only
3. The 2017 to 2018 sample for CT small business was not used in 'Measuring tax gaps 2020' due to operational issues
H.5 To produce population estimates for total tax gaps from the samples in Table H.1, the average tax gap estimates from random enquiries are multiplied by the number of taxpayers in the population.

H.6 Adjustments are made to the population for cases deselected because they are outside of the population of interest, for example, the business is no longer operating or is part of the large business customer group.

**Self Assessment**

H.7 The Self Assessment (SA) random enquiry programme allows us to estimate the tax gap arising from under-declaration of tax liabilities of individuals in SA and is used in conjunction with the operational enquiry data. Results from enquiries are scaled up to the total number of individuals who sent a SA notice to file and who are in the population of interest. For individuals not covered by the random enquiry programme, operational enquiry data is used. Further details regarding this data are given in the ‘Enquiry Data’ section of this chapter.

H.8 In this context, ‘individuals’ means individuals who are self-employed, pensioners, and partnerships (with up to four partners), as well as those who are employees or may only have investment income. The taxes directly included are:

- IT
- NICs
- CGT

H.9 The random sample used for the programme is selected from SA taxpayers issued with a notice to file a return. The sample is drawn by a systematic process that selects every \(^n\)th notice. The sampling interval, \(n\), is determined by dividing the total number of returns issued by the required sample size (rounded down to the nearest whole number). When the return includes a partnership income schedule, we deselect that return. This is because the returns of individuals who are partners will automatically be included in any enquiry resulting from the selection of a partnership return.

H.10 2009 to 2010 is the last year which uses a simple random sample, as random samples for subsequent years have been stratified to improve the accuracy of the results. Samples drawn from SA business taxpayers are stratified by turnover from 2010 to 2011 onwards, with samples drawn from SA non-business taxpayers stratified by level of income from 2011 to 2012 onwards.

H.11 From 2015 to 2016 we used an optimal allocation method in order to increase the accuracy of our estimates. When sampling, we take into account the variability of the tax at risk across the strata in the population. We select a greater proportion of cases in strata where the variance of tax at risk values is known to be high.

H.12 SA business consists of the self-employed and partnerships. SA non-business consists of employees, pensioners, trusts and all other types of SA taxpayers. In order to improve how representative the sample is, a weighting is applied based on how these customer groups are distributed across the population. We continue to review the customer group population assumptions.

H.13 Due to a relatively small sample size and large natural variance in the levels of under-declared liabilities from year to year, a smoothing approach has been used for small partnerships from 2010 to 2011, when the stratification of business taxpayers was introduced. A three-year moving average with a double weighting given to the current year is used to smooth the data. This ensures that the resulting estimates are less susceptible to sampling variability and more indicative of longer-term trends.
**Employer compliance**

H.14 The employer compliance (EC) random enquiry programme allows us to estimate the tax gap arising from Pay As You Earn (PAYE) failures and other irregularities. Results from the EC random enquiry programme are scaled up to the total number of PAYE schemes.

H.15 The employer may be a self-employed individual, partnership, or a company and will be required to make returns under the PAYE regulations to account for IT and NICs.

H.16 The figures relate solely to IT and NICs collected through PAYE due on earnings and other income from employment. The scope of these figures also includes tax due on occupational pensions taxed through PAYE.

H.17 The taxes directly included are:

- IT
- NICs
- tax on occupational pensions

H.18 The random sample is selected using the former small and medium-sized enterprises (SME) customer classification and stratified on the basis of employers segments (defined in terms of the number of employees and whether the employer’s business is incorporated). Prior to ‘Measuring tax gaps 2020’ we calculated the small business tax gap by first calculating the EC tax gap for SME employers. This SME estimate was then converted to a small business estimate using historical data on tax receipts that was available under both groupings. From ‘Measuring tax gaps 2020’ the EC small business tax gap is calculated directing using small business only data for the tax year 2015 to 2016 and onwards. This is done by flagging and removing mid-sized businesses from the random sample to create a small business only sample. However, for historic years the conversion factor is still applied as the source data does not contain the information required for the identification of mid-sized businesses, therefore they cannot be removed.

**Corporation Tax**

H.19 The CT random enquiry programme allows us to estimate the tax gap arising from incorrect CT returns of small businesses. Results from the CT random enquiry programme are scaled up to the total number of live small business trader cases. In this context, ‘live’ excludes cases which are, for instance, dormant or dissolved/struck off.

H.20 For CT, up to the tax year 2015 to 2016, the random sample was selected using the former SME customer classification. From 2016 to 2017 the random sample is selected from the small business customer group from businesses which have been issued a notice to file a return. For the ‘Measuring tax gaps 2020’ edition the results of the 2017 to 2018 random sample were not used to estimate the CT tax gap. This is due to the results from the 2017 to 2018 sample not being available in time for this year's publication.

H.21 The random enquiry is used to directly calculate the CT small business tax gap from the tax year 2016 to 2017 and onwards. For earlier years the random enquiry data was collected under the former SME population definition and the estimate of the small business tax gap was estimated by applying a conversion factor which was derived using historical data on tax receipts that was available under both groupings.

H.22 From April 2013, the sampling process changed to a stratified random sample, based on the size of annual trading turnover. This change allowed the CT random enquiry results to be weighted by the actual population of each strata resulting in an improved accuracy of the tax gap results. In addition to this, cases are excluded if the company has not submitted a return for the year of interest.
Due to a relatively small sample size and large natural variance in the levels of under-declared liabilities from year to year a smoothing approach is used. A three-year moving average with a double weighting given to the current year is used to smooth the CT small business data throughout the series. This ensures that the resulting estimates are less susceptible to sampling variability and more indicative of longer-term trends.

**Data features**

### H.24

The latest observed random sample for SA used in the ‘Measuring tax gaps 2020’ edition estimates are for 2016 to 2017. More detail of the timing of random enquiries is given in the next section. From 2014 to 2015 approximately half of the sample was worked as a desk-based enquiry rather than the standard face to face approach before the move to a fully desk-based approach was implemented in the tax year 2016 to 2017. An evaluation of the effect of working REP cases as a desk-based enquiry as opposed to face to face was carried out and found no statistically significant evidence that it affected the outcome of the enquiry.

### H.25

The latest observed EC random sample is for 2018 to 2019. From 2015 to 2016, approximately half of the sample was worked as a desk-based enquiry rather than a face to face approach before the move to a fully desk-based approach was implemented in the year 2018 to 2019. An evaluation of the effect of working cases as a desk-based enquiry as opposed to face to face was carried out and found no statistically significant evidence that it affected the outcome of the enquiry.

### H.26

The latest year of CT REP data relates to 2017 to 2018, however we were unable to use this data in our estimates. This is due to the results from the 2017 to 2018 sample not being available in time for this year's publication. From 2017 to 2018 approximately half of the sample was worked as a desk-based enquiry rather than the standard face to face approach. We will evaluate the impact of this change ahead of ‘Measuring tax gaps 2021’.

### Timing

### H.27

There are two factors which influence the timing of the latest available tax gap estimate for a particular type of tax return:

- delays inherent in the returns process; this varies according to the head of duty and is shown in Table H.2 below
- delays due to the complexity of some random enquiries; it can take several years before sufficient random enquiries relating to a particular tax year are settled to robustly report the results
Table H.2: Comparison of delays due to returns process

<table>
<thead>
<tr>
<th>Random enquiry programme</th>
<th>Delays due to returns process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self Assessment</td>
<td>Individuals generally have until 31 January following the year of assessment to which the return relates to submit their return. Once the return is submitted, HMRC then has a further year in which to open an enquiry.</td>
</tr>
<tr>
<td>Employer compliance</td>
<td>None. EC reviews initially look at the records of the previous 12 months.</td>
</tr>
<tr>
<td>Corporation Tax</td>
<td>Companies have until a year after the end of their accounting period to submit their return. HMRC then has a further year in which to open an enquiry.</td>
</tr>
</tbody>
</table>

H.28 There are consequences of the timing issues described above. Firstly, estimates of tax gaps for CT and SA are not available for the latest years due to a lag in data available so to present a more consistent picture of the scale of tax losses, projection factors have been applied to the estimates for CT and SA. We use the latest available data to project future years as this allows us to most effectively reflect recent policy and other changes that have a long-term impact on taxpayer behaviour. These projection factors are shown below in Table H.3.

H.29 Secondly, estimates for earlier years have been revised since previously published, as a result of the inclusion of additional data from reviews that have since been completed. Finally, at the time of estimation, some enquiries were not closed for each year of each random enquiry programme. In order to estimate tax gaps for each year, it is necessary to make assumptions about the cases that were yet to be settled at the date the enquiry results are analysed. Forecasts for such enquiries are made based on the results of recently settled enquiries with similar durations.

Table H.3: Comparison of projection factors

<table>
<thead>
<tr>
<th>Random enquiry programme</th>
<th>Projection factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self Assessment</td>
<td>For the years after 2016 to 2017 where we do not have random enquiry data, the latest available estimate is projected forward. The projections are made by keeping the percentage gross tax gap constant and using actual tax liabilities, non-payment and compliance yield for the relevant tax.</td>
</tr>
<tr>
<td>Corporation Tax</td>
<td>For the years after 2016 to 2017 where we do not have random enquiry data, the latest available estimate is projected forward. The projections are made by keeping the percentage gross tax gap constant and using actual tax liabilities, non-payment and compliance yield for the relevant tax.</td>
</tr>
<tr>
<td>Employer Compliance</td>
<td>No projection factor is used.</td>
</tr>
</tbody>
</table>
Sources of error

H.30 There are two main sources of error associated with the results of random enquiries which could result in the true values of the tax gaps differing from the estimates produced. These are:

- sampling variation in the data: the whole population is not subject to enquiry, so even though the sample is designed to be representative, its characteristics may differ from the population purely by chance

- systematic uncertainty where the sample results consistently tend to under-report the true values for the population, or where the sample does not include the full population, for example those participating in avoidance

H.31 We make an adjustment for one source of systematic uncertainty, which is non-detection of non-compliance. The random enquiry programmes will not identify all incorrect returns or the full scale of under-declaration of liabilities, and so estimates produced from the unadjusted results of the programmes would underestimate the full extent of the tax gap. The Internal Revenue Service (IRS) in the United States (US) has previously tackled this problem by using a range of ‘multipliers’ to adjust for non-detection. The principles behind the IRS methodology have been applied to HMRC’s data to produce approximate multipliers for the UK. To read the IRS report go to [www/darp.lse.ac.uk/papersdb/Andreoni_etal_%28JEL98%29.pdf](http://www/darp.lse.ac.uk/papersdb/Andreoni_etal_%28JEL98%29.pdf).

H.32 The IRS was able to undertake this analysis of non-detection because their random enquiry samples covered upward of 50,000 cases – much higher than is feasible in the UK. In the absence of this data for the UK, the US multipliers are used to account for non-detection. The size of the multipliers varies by the type of non-compliance found and are consistent year-on-year; Table H.4 shows how these multipliers differ by each random enquiry programme.

<table>
<thead>
<tr>
<th>Random enquiry programme</th>
<th>Multiplier for central estimate</th>
<th>Multiplier for lower estimate</th>
<th>Multiplier for upper estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self Assessment (business)</td>
<td>1.908</td>
<td>1.000</td>
<td>3.075</td>
</tr>
<tr>
<td>Self Assessment (non-business)</td>
<td>1.260</td>
<td>1.000</td>
<td>1.928</td>
</tr>
<tr>
<td>Employer compliance</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Corporation Tax</td>
<td>1.376</td>
<td>1.000</td>
<td>1.859</td>
</tr>
</tbody>
</table>

Modelling adjustments

H.33 When calculating the SA tax gap as a percentage of liabilities in previous publications, we amended the liability values to account for forestalling. This meant that we adjusted the liability values to undo changes in taxpayer behaviour caused by the announcement of future tax policy changes.

H.34 For ‘Measuring tax gaps 2020’ we have stopped making the adjustment to liabilities to account for forestalling. This is because we want the tax gap to reflect actual taxpayer behaviour rather than the behaviour that would have been observed if there had been no announcements of changes to tax policy.
Validation

H.35 As part of each year’s programme, HMRC conducts a validation exercise for a sample of cases. These cases are checked to confirm that the enquiry outcomes (for example, the amount of yield) have been recorded accurately. Any inaccuracies are corrected prior to calculation of the tax gap for that year. Work is underway on how best to use the results of this exercise to allow the correction of systematic errors to be projected onto the rest of the sample in a statistically valid way.

Outliers

H.36 Outliers are individual cases with large yields which are far removed from the yields of the other cases in the sample. Due to the nature of our samples our estimates are particularly sensitive to extreme values. To ensure that this small number of cases do not have an undue influence on the tax gap calculation their yield values are therefore capped. This allows us to use all valid information while smoothing the year-on-year variability.

H.37 Yield data is modelled using a representative statistical distribution. The final value used for each tax year is calculated as a 3-year moving average of the 99.85th percentile from this distribution, calculated based only on the results of years where the sample was stratified. For years before stratification, and years where a full 3 years of stratified results are not available, the value based on the last 3 complete stratified years is used.

H.38 A specific capping value is calculated for each random enquiry programme, including a separate value for SA business and non-business.

Deselections

H.39 Cases in the random enquiry programme are not worked for a number of reasons and this is done in a non-random way. This means that the cases which are not worked are likely to be systematically different from the cases that are worked. Cases which are not worked are called deselections or rejections depending at which stage of the production process the decision to not work the case was taken. To avoid biasing the sample we treat and include cases that are deselected from the sample but are still within the population of interest. If the individual or business has undergone a recent enquiry, we substitute the outcome of this earlier enquiry into the case. If no such previous enquiry exists, we assign a value based on the average yield and probability of being non-compliant in the taxpayer’s stratum.

Enquiry data

H.40 HMRC operates a comprehensive system of targeted customer auditing that includes monitoring, carrying out risk assessments, and from this making resourcing decisions to better direct enquiries towards the highest risk customers.

Data Issues

H.41 There is a lag in the data available for the most recent years in self-assessment inherent to the returns process. See paragraph H.26 for more details. This has two main consequences:

- estimates of tax gaps for SA are not available for the latest years due to the lag in data available so to present a more consistent picture of the scale of tax losses, projection factors have been applied to the estimates for the SA enquiry data
• at the time of estimation, some enquiries were not closed so to estimate tax gaps for each year, it is necessary to make assumptions about the cases that were yet to be settled at the date the enquiry results are analysed

H.42 In order to maintain the integrity of the time series, we have created an illustrative estimate of the tax gap based on the total SA tax gap for the years prior to 2015 to 2016. This is because the operational enquiry data is not in the required format in earlier years due to operational reasons.

H.43 The cases will not identify all incorrect returns or the full scale of under-declaration of liabilities, and so estimates produced from the unadjusted results of the enquiry data would underestimate the full extent of the tax gap. We apply a non-detection multiplier to get a better reflection of what the true tax gap would look like. See paragraph H.31 for more details.

H.44 The Delphi technique was used to derive specific non-detection multipliers for the operational enquiry data. We use different non-detection multiplier depending on the perceived risk of the cases ranging from 1.5 to 1.7. More information about the Delphi approach and non-detection multipliers can be found in HMRC’s working paper [website link].

Methodology

H.45 For cases which were audited, we sum the yields and apply a non-detection multiplier to get the gross tax gap. Unaudited cases are estimated by creating an upper and lower bound by scaling the results from the highest and lowest risk segments to the population size and then adjusted by a non-detection multiplier. The gross tax gap estimate from the enquiry data is then the sum of the tax gap from audited cases and the midpoint between the upper and lower bounds for unaudited cases.

H.46 There is no direct way of splitting the enquiry data between SA business, SA non-business and SA large partnerships. We split the tax gap between these groups by looking at the percentage of the population that falls into each of these groups and applying these percentages to the tax gap.

Tax gap calculation

H.47 The methodology for the EC and CT small business tax gaps combines the estimate of under-declared liabilities with the amount of non-payment. As some of the tax gap is recovered through HMRC compliance activity, this is subtracted to give the net tax gap.

H.48 The flowchart below illustrates the series of model operations described above using symbols to represent each step of the process and contains a short description of the process step to calculate the EC and CT tax gap estimate:

Net tax gap = \( \text{Under-declared liabilities from incorrect returns} \times \text{US multipliers to account for non-detection} \) + \text{Non-payment} - \text{Yield from compliance activity}

H.49 The methodology for the SA tax gap uses the combined estimate of undeclared liabilities from the random enquiry program and the enquiry data. As for CT and EC small business non-payment is added on to the total estimate of under-declared liabilities and yield from compliance activity is subtracted to give the net tax gap.
The flowchart below illustrates the series of model operations described above using symbols to represent each step of the process and contains a short description of the process step to calculate the SA tax gap estimate:

\[
\text{Net tax gap} = (\text{Under-declared liabilities from REP} \times \text{US multipliers to account for non-detection})
+ (\text{Under-declared liabilities from enquiry data} \times \text{Delphi multipliers for non-detection})
+ \text{Non-payment} - \text{Yield from compliance activity}
\]

The ranges which define the upper and lower estimates of the tax gap are based on the 95% confidence intervals of the estimate for under-declared liabilities from incorrect returns. These ranges are adjusted for non-detection as described in Table H.4 above.

**Non-payment**

The figures used to estimate levels of non-payment come from analysis of write-offs and remissions of tax on a financial year basis.

As separate figures of non-payment are not available for just the taxpayers within the scope of the random enquiry programmes, the amounts are split in proportion to the tax gap resulting from the relevant section of the populations. These non-payment figures will relate to the year when the loss was realised rather than the tax year the liability relates to. This approach has been taken because figures are not readily available by reference to the liability period.

**Compliance yield**

The random enquiries provide an estimate of the tax gap due to incorrect returns. However, HMRC carries out a wider programme of compliance activity to identify and correct erroneous returns. To calculate the net tax gap, it is necessary to subtract the yield from this activity. The figures for yield are taken from HMRC’s systems for recording the outcomes of enquiries and relate to cases settled during each year rather than enquiries into returns relating to a specific tax year. See Chapter B Accuracy and reliability.

**Other estimates**

Large employers operating a PAYE scheme

Larger employers are not covered by the EC random enquiry programme and we intend to undertake further methodological development to produce a robust estimate of this gap. We will continue to review our methods, including the option of using risk-based models as described in Chapter I, for estimating the large business EC gap in future.

An illustrative estimate is produced by assuming that the tax at risk will represent, over the long term, the same proportion of liabilities to small and the smallest of mid-sized employers, as shown by the results of the random programme. The estimated tax at risk is then adjusted to reflect compliance yield and non-payment.

An adjustment to the estimate of the tax gap was made following on from the introduction of the Real Time Information (RTI) system, where information on payroll taxes is recorded more accurately and on a more frequent basis allowing us to identify debts and take action at an earlier stage than previously. This is done by estimating the impact of RTI on the tax gap estimates from the random enquiry programme and applying this change to the estimate for the larger employers.
Large partnerships in Self Assessment

H.58 An illustrative estimate has been produced by assuming that the tax at risk will represent a similar proportion of liabilities to all other SA taxpayers, as shown by the results of the SA random enquiry programme. Projections for 2017 to 2018 and 2018 to 2019 are made by keeping the percentage gross tax gap constant and using actual tax liabilities, non-payment and compliance yield.

Wealthy

H.59 To calculate the SA wealthy tax gap, we use operational enquiry data to match the wealthy population to both the enquiry data and the REP data. We then apply the same methodology used to estimate the SA tax gap to the wealthy population to get the wealthy portion of SA business and SA non-business tax gaps.

H.60 The large partnerships tax gap is not measured directly and as such we cannot use data matching in order to find the wealthy portion of the tax gap. Instead, we find the percentage of liabilities from large partnerships which comes from wealthy taxpayers and apply that percentage to the SA large partnership tax gap.

H.61 The net tax gap is calculated by taking the compliance yield away from the gross tax gap and adding on non-payment.
Chapter I: Estimates using risk-based enquiry programmes

I.1 For ‘Measuring tax gaps 2020’ edition we have introduced methodologies which utilise HMRC’s risk-based operational enquiry data to estimate areas of the tax gap. These statistical methods have been used to produce the Corporation Tax (CT) gap for large and mid-sized businesses, the Employer Compliance (EC) gap for mid-sized businesses, as well as the Inheritance Tax (IhT) gap.

I.2 Previously the mid-sized business EC, large and mid-sized business CT and IhT gap estimates were all illustrative. Using the methods described in this chapter, we now use data specific for these customer groups in our estimates. This means the scale of the gap is likely to be a far better estimate of the true non-compliance in the respective populations. It also means the gap will be dynamic and reflect real changes in compliance over time.

I.3 The EC gap estimates for large businesses remain illustrative at this time and are described in Chapter H.

Risk-based estimates

I.4 HMRC carry out risk assessments to determine when and how to enquire into cases where there is a risk that the taxpayer has not paid the correct tax. These enquiries can take many different forms, such as a Self Assessment tax enquiry, an Employer Compliance review, or a VAT audit.

I.5 Unlike random enquiry programmes (see Chapter H) these enquiries are not representative of the population and require statistical methods to extrapolate the results to the unaudited population as part of a tax gap calculation.

I.6 There are three taxes where we have utilised risk-based enquiries to produce tax gap estimates. They are:

- Employer Compliance (EC) for mid-sized business employers
- Corporation Tax (CT) for mid-sized and large businesses
- Inheritance Tax (IhT)

Extreme value methodology

I.7 Extreme value (EV) methodology is a statistical technique that is used to understand data that is characterised by extreme outlier observations. An example of this would be where a small number of data points make up a large majority of the total value. This is known as the power law or 80-20 rule, where approximately 80% of the yield comes from 20% of enquired cases.


I.9 Operational enquiry results from the populations studied in this chapter fit the extreme values description – we find that most of the non-compliance is concentrated in a small number of businesses/individuals.
Data lying outside the extreme values regime fails to follow the law and is removed by the implementation of a threshold cut-off. The remaining data is fitted to the power law model using an ordinary least squares (OLS) regression.

The number of unaudited businesses/individuals that might contribute to the tax gap is calculated by assuming the same proportion of above threshold cases as found in the audited population. These businesses, having not been chosen for an enquiry, are assumed to be less risky and unlikely to contribute large tax adjustments.

Tax gap calculation

The higher the accuracy of the risking methods the more accurate the EV results will be. We have taken a conservative assumption that the EV method may lead to an underestimate of the tax gap in some of our population groups where the risking procedures are more challenging and some higher yield cases may have been missed.

Employer Compliance & Corporation tax

In the case of the EC gap for mid-sized employers and the CT gap for mid-sized and large employers we combine the results of the EV methodology with an upper bound estimate.

For the upper bound, the average tax gap, as a percentage of liabilities, for the unaudited businesses is assumed to be the same as that for audited businesses. This leads to an overestimate since the audited businesses have been selectively chosen for audit on the basis of their expected non-compliance.

The gross tax gap for these populations is then found as the average of the EV and upper bound results.

Inheritance tax

In the case of the IhT gap, the EV is the central estimate due to comprehensive risking of the tax paying IhT return, the IHT400. An adjustment is made for Lifetime IhT, which is liabilities arising on immediately chargeable transfers during life and on trustees on assets held in trusts. Since suitable operational audit data for the Lifetime IhT is not available, a part of the tax gap is assumption based. The gross tax gap for IhT is therefore the sum of the EV and lifetime estimate.

The net tax gap is calculated by taking the compliance yield away from the gross tax gap.

Data Issues

Operational audit data in a suitable form is not available for some earlier years. As such the tax gaps that have adopted this methodology have done so from the tax years 2013 to 2014 for IhT, and 2014 to 2015 onwards for the CT and EC gaps.

The earlier years have been calculated by scaling the previously published data by the ratio between the old and new results, in the years both figures were available.

Identified risks can take many years to resolve. For open enquiries it is necessary to forecast the expected compliance yield to calculate the tax gap. We do this by matching all open cases to a similar closed case and using the closed case yield in the model. Differences between the forecast yield and actual yield may lead to revised tax gap estimates in subsequent publications, but the use of forecasting reduces the chance that these revisions are significant. The tax gap for more recent years is likely to be subject to larger revisions because a higher proportion of the compliance yield is estimated.
I.21 Risks may also take several years to identify, and this is significant in the data for CT for more recent accounting periods. The use of projected data for these years ensures the chance of large revisions to these years is minimised.

I.22 Risks can relate to multiple accounting periods. The yield collected is allocated evenly between these periods in the absence of more detailed information to break it down. Risks can also be identified and not lead to any tax adjustment if the taxpayer is deemed to be compliant. These are still included in our model as they are an important contribution to the overall picture of (non)-compliance in the population.

Sources of error

I.23 There are two main sources of error associated with the results of these methodologies which could result in the true values of the tax gaps differing from the estimates produced. These are:

- systematic uncertainty where the results from the risk-based audits consistently tend to under-report the true values for the population, or where the sample does not include the full population, for example those participating in avoidance
- uncertainty due to variations in the risk-based data due to different risking approaches
- Inaccurate population numbers

I.24 We make an adjustment for one source of systematic uncertainty, which is non-detection of non-compliance. Audits will not identify all incorrect returns or the full scale of under-declaration of liabilities, and so estimates produced from the unadjusted results of the programmes would underestimate the full extent of the tax gap. To account for this, we apply a non-detection multiplier to audit data to get a better reflection of what the true tax gap would look like.

I.25 For the CT and EC models we use multipliers that are the same as those for the Small customer group and further details can be found in Chapter H ‘Estimates using random enquiry programmes’.

I.26 For Iht, we base the multiplier on expert opinion. This involved HMRC operational colleagues reaching a consensus through a short questionnaire, which was further quality assured by a wider stakeholder group. These non-detection multipliers are subject to revision as we are planning to replace these with a new methodology. More information can be found in HMRC’s working paper www.gov.uk/government/publications/non-detection-multipliers-for-measuring-tax-gaps.

I.27 The non-detection multipliers currently used in the EV methodologies are shown below:

<table>
<thead>
<tr>
<th>Non-detection multiplier</th>
<th>Corporation Tax</th>
<th>Employer Compliance</th>
<th>Inheritance Tax</th>
</tr>
</thead>
</table>

|                | 1.376 | 1.000   | 1.700 |

I.28 Data variations arising from changes to the success of the risk profiling used to obtain operational audit data has the potential to lead to changes in the tax gap estimate that may not be reflective of the real-world situation. An example would be a change to the success of the risking procedures.

I.29 It is challenging to quantitively measure risking success, but we gain an understanding of this through discussions with business experts and future adjustments to the model, in particular the use of the upper bound method and the proportions of the model we take from this.
I.30 Population numbers are fairly static and well defined in the case of mid-sized and large business. The IHMT population isn’t well defined, but it is approximated using the total number of estates above the nil-rate band. Sensitivity analysis has been carried out that confirms that the population number is not a very sensitive input.

Non-payment

I.31 The figures used to estimate levels of non-payment come from analysis of write-offs and remissions of tax on a financial year basis.

I.32 Non-payment figures will relate to the year when the loss was realised rather than the tax year the liability relates to. This approach has been taken because figures are not readily available by reference to the liability period.

Compliance yield

I.33 The compliance yield for risk-based models is calculated as the total yield from closed risks plus the estimated compliance yield from open risks. Compliance yield in these models relates to a specific accounting period and therefore cannot be compared to business reported compliance yield.
Chapter J: Other taxes

J.1 Other taxes include:

- Other direct taxes
  - Inheritance Tax (this is described in chapter I)
  - Stamp Duty Land Tax
  - Stamp Duty Reserve Tax
  - Petroleum Revenue Tax

- Other indirect taxes:
  - Aggregates Levy
  - Air Passenger Duty
  - Betting and gaming duties
  - Climate Change Levy
  - Customs Duty
  - Insurance Premium Tax
  - Landfill Tax
  - Soft Drinks Industry Levy (introduced in tax year 2018 to 2019)
  - Spirit-based ready-to-drink duties
  - Still cider and perry duties
  - Wine duty

J.2 With the exception of Stamp Duty Land Tax and Inheritance Tax, methodologies for ‘other taxes’ are experimental where we use the best available data, simple models and management assumptions to build an illustrative estimate of the tax gap.

J.3 Petroleum Revenue Tax is only estimated up until the 2014 to 2015 tax year. After this, the estimate discontinued due to Petroleum Revenue Tax being permanently zero-rated from 1 January 2018.

Stamp Duty Land Tax

Methodology

J.4 The Stamp Duty Land Tax (SDLT) tax gap is an established methodology and is estimated using a combination of management information and management assumptions.

J.5 The tax at risk (TAR) from avoidance cases being investigated by HMRC was projected this year using the 2017 to 2018 data.

Tax under consideration

J.6 The SDLT tax gap is calculated from the amount of SDLT outstanding, referred to here as tax at risk. Four components which contribute to the tax gap have been identified:

- TAR from cases being investigated by the SDLT Counter-Avoidance team
- SDLT avoidance unknown to the department
- reliefs that are improperly claimed
• SDLT not paid due to evasion, goodwill, agent behaviour and linked transactions

**SDLT avoidance unknown to the department**

J.7 It would be impossible for HMRC to know about every case of SDLT avoidance, because either the associated paperwork has not been completed, or because it has been deliberately falsified and not yet discovered, or for some other reason. Expert opinion has suggested that HMRC is likely to be aware of approximately 80% of all transactions involving SDLT where tax at risk has resulted. For this reason, a multiplier of 1.25 (100 / 80) has been used to 'uplift' the amount of known tax at risk to account for this.

**Evasion**

J.8 This reflects a percentage of the total amount of SDLT receipts (as published by HMRC) not initially paid because of evasion. Internal discussions with subject matter experts suggested that this amounts to 1% of the published SDLT receipts each year, with around 50% of this recoverable in line with other non-avoidance activity.

**Reliefs improperly claimed**

J.9 Improperly claimed relief takes different forms and there are more than 30 different reliefs claimed for SDLT. All reliefs are taken into account for this calculation.

J.10 Analysis of open enquiries and a series of pilot research projects have suggested that up to 5% of these claims may be falsely claimed. Additionally, there is an assumption that HMRC may only be able to recover 10% of the tax at risk involved in these cases: this takes into account the large number of reliefs for which compliance work has not yet begun and the small number of cases open into those reliefs that have been targeted.

**Goodwill, agent behaviour and linked transactions**

J.11 This reflects a percentage of the total amount of SDLT receipts (as published by HMRC) not initially paid because of goodwill, agent behaviour and linked transactions. Internal discussions with subject matter experts suggested that this amounts to 0.5% of the published SDLT receipts each year, with around 50% of this recoverable in line with other non-avoidance activity.

**Exclusions from this methodology**

J.12 Estimates for years prior to 2011 to 2012 include the amount of SDLT avoided by the use of tax avoidance schemes. These were artificial structures solely constructed to avoid SDLT that the department was aware of. This was calculated by multiplying together the number of disclosures of tax avoidance schemes (DOTAS) schemes, the estimated tax under consideration each year and the estimated number of users of each DOTAS scheme. This is excluded from 2011 to 2012 onwards as it is believed that no further DOTAS schemes related to SDLT have been revealed to the department.

J.13 Estimates for years prior to 2015 to 2016 include threshold manipulation (another form of SDLT evasion). This occurred when a sale value of a property was artificially reduced to below a threshold in order to reduce the SDLT liability. Previously, SDLT was charged at a single rate based on the value of the total purchase price. From 4 December 2014, SDLT liabilities changed to incremental rates applied only to the portion of the purchase price that falls within each rate band. This significantly reduced the potential value of tax lost due to threshold manipulation. For this reason, estimates after this point do not include threshold manipulation.
Landfill Tax

Methodology

J.14 The Landfill Tax gap is estimated using an experimental methodology using a combination of modelling, proxy indicators and assumptions made in collaboration with HMRC’s operational experts. It uses HMRC, Environment Agency (EA) and publicly available data to estimate each component.

J.15 From 1 April 2015, Landfill Tax was devolved to Scotland hence, since ‘Measuring tax gaps 2017’ edition, Scottish Landfill Tax is no longer in scope of this estimate. Landfill Tax attributable to Scotland is removed from the tax gap estimate by using the percentage of total UK Landfill Tax receipts attributable to Scotland.

J.16 From 1 April 2018, Landfill Tax was devolved to Wales hence, since ‘Measuring tax gaps 2020’ edition, Welsh Landfill Tax is no longer in scope of this estimate. Landfill Tax attributable to Wales is removed from the tax gap estimate by using the percentage of total UK Landfill Tax receipts attributable to Wales.

Tax in scope

J.17 Landfill Tax is due on waste disposed of at permitted landfill sites and at unauthorised waste sites as a disincentive to landfilling and to encourage better waste management. The tax gap measures the difference between the amount of Landfill Tax that should theoretically be collected, when waste is disposed of at a permitted landfill site and at unauthorised waste site, and the amount that is actually collected.

J.18 The methodology has been updated this year to include unauthorised waste sites. This became taxable from 1 April 2018 so will be included in publications for tax years 2018 to 2019 onwards. Tax was not due on unauthorised waste sites prior to this date, so has been excluded from our methodology for earlier years up to and including 2017 to 2018.

Tax under consideration - under-declaration

J.19 Under-declared waste is estimated in two ways and averaged to arrive at a central estimate.

J.20 In the first method a trend line is fitted to HMRC data on taxable tonnes over time, then expected and actual tonnages of waste are compared. The estimate is refined to take account of the increase in diversion of waste away from landfill in recent years to incineration and export of refuse derived fuel. We assume nearly all of this diverted waste is taxable at the standard rate if sent to landfill.

J.21 After these adjustments, the tax under consideration is estimated by applying the tax rates at the same composition as declared taxable waste. The ratio of standard rate to lower rate has changed over time with it becoming roughly 50:50 in recent years.

J.22 In the second method, a proxy indicator is used to estimate under-declaration. This assumes that all landfill site operators have under-declared taxable waste by 5% per year, and that this under-declared amount should be taxed at the standard rate.

Tax under consideration - misclassification

J.23 There are two rates of Landfill Tax, standard and lower rate. A trend line is fitted to HMRC published statistics on lower rated tonnes declared over time. Expected tonnages of lower rate waste is then compared with declared lower rate waste. Declared lower rate waste shows a trend towards increasingly larger amounts of lower rated waste going to landfill in recent years. Some of this is expected due to changes in how waste is diverted away from landfill towards other forms of waste management.
J.24 We assume 25% of the difference between expected and declared lower rated waste constitutes the tax base under consideration. The tax under consideration is then the difference between the standard and lower rates of waste on this tonnage.

**Unauthorised waste sites**

J.25 The estimated tax gap is based on EA data on estimated tonnage for known illegal waste sites. This data is provisional while the EA undertake a quality review of the estimated tonnage for the largest sites, which may result in revisions to future estimates.

J.26 Sites that gained exemption or the appropriate permit during the tax year 2018 to 2019 have been excluded from our calculation, as have sites that have evidence of being stopped, regulated or cleared before 1 April 2018. The standard rate of Landfill Tax has been applied to the remaining sites to calculate the tax gap for England. The estimate is uplifted to account for Northern Ireland (NI), which isn’t covered by EA data. This is done by applying the ratio of total tax liabilities in England and NI to the estimate for unauthorised waste sites tax gap in England.

**Tax gap calculation**

J.27 As some of the gross tax gap is recovered through HMRC compliance activity, this is subtracted to give the net tax gap. The net tax gap estimate is defined as:

\[
\text{Net tax gap} = \text{Under-declared waste} + \text{Misclassified waste (as either standard rated or lower rated)} + \text{Unauthorised waste sites} - \text{Yield from compliance activity}
\]
Chapter K: Avoidance and hidden economy

Avoidance

Data sources

K.1 This section describes estimates of the avoidance tax gap for IT, NICs and CGT. The avoidance tax gap is estimated using information that HMRC collects on tax avoidance schemes and records on its management information system. This includes avoidance schemes for individuals, trusts, partnerships and employers. The information that HMRC collects relates to ‘disclosed’ and ‘undisclosed’ schemes.

K.2 ‘Disclosed’ schemes are arrangements (including any scheme, transaction or series of transactions) that will or are intended to provide the user with a tax advantage when compared to a different course of action and, under tax legislation, must be disclosed to HMRC. For more information about disclosure of tax avoidance schemes (DOTAS), go to www.hmrc.gov.uk/aiu/summary-disclosure-rules.htm.

K.3 ‘Undisclosed’ schemes, are arrangements identified by HMRC, not disclosed under DOTAS legislation.

K.4 For schemes disclosed under DOTAS, information is captured during the following process:

- promoters of avoidance schemes that are covered by the avoidance disclosure rules must disclose any new schemes to HMRC when they are made available to potential users
- disclosures must contain sufficient detail for HMRC tax specialists to understand how the scheme works
- for each disclosure HMRC issues a scheme reference number to the promoters, and taxpayers who participate in the scheme are required to notify HMRC of the reference number on their tax return (described here as a ‘notification’)

K.5 When reviewing both ‘disclosed’ and ‘undisclosed’ avoidance schemes, tax specialists record an estimate of the ‘tax under consideration’ based on the relevant information relating to these ongoing enquiries. Any additional tax (‘compliance yield’) that is collected following completed enquiries is also recorded.

K.6 Detailed taxpayer-level data on avoidance schemes is available for large businesses and wealthy individuals. This enables comparison of the tax under consideration and compliance yield for an individual scheme user. Data on completed enquiries provides a basis to estimate expected compliance yield from ongoing enquiries.

Methodology

K.7 The tax gap is calculated by subtracting estimated compliance yield from tax under consideration:

\[
\text{Estimated tax gap} = \text{Tax under consideration} - \text{Estimated compliance yield}
\]

K.8 The tax under consideration estimate relates to ongoing and completed enquiries. For completed enquiries, an estimate of tax under consideration is calculated from the compliance yield figures. This is calculated by applying the ratio of the compliance yield to tax under consideration from the taxpayer-level data to the actual compliance yield data.
K.9 The compliance yield that is likely to be recovered for those under investigation is estimated using the ratio of the compliance yield to tax under consideration. This ratio is derived from the taxpayer-level data on completed avoidance enquiries.

Data quality

K.10 The main source of error in these estimates is that HMRC may not identify all avoidance schemes – which will lead to an underestimation of the tax gap. It is difficult to quantify the extent to which this source of error impacts upon the estimates.

K.11 There are a number of issues with the methodology to estimate the avoidance tax gap:

• estimates of tax under consideration are made by tax specialists using all the information available at the time; as this information improves over time, the view of tax under consideration may change

• the ratio of compliance yield to tax under consideration will change over time as more enquiries are completed; any difference between estimated compliance yield from ongoing enquiries and actual compliance yield will lead to revisions in the estimates

• there is no tax year attached to the ‘tax under consideration’; therefore, the distribution of scheme uses across tax years is used to derive an annualised estimate

• CT avoidance for LB groups are excluded from the calculations to avoid double-counting with the separate avoidance estimate for these businesses; any re-classification of users following more accurate information would lead to revisions of the CT avoidance estimate

K.12 As a result of these factors, the figures presented in the document are likely to be revised as more information becomes available.

K.13 The data on avoidance schemes are reviewed by HMRC analysts for consistency and accuracy. Over time, as the scope, quality and quantity of the data improves, HMRC will seek to improve the avoidance tax gap estimates.

K.14 For ‘Measuring tax gaps 2020’ edition the 2018 to 2019 estimate is a projection from the 2017 to 2018 due to issues with data processing and we anticipate revisions in future years. The issue meant that while we could determine where HMRC had identified avoidance schemes and how many individuals had used them, we could not allocate the most recently available data to the correct tax years within our avoidance tax gap analytical model. We will revise our avoidance estimate in future editions of ‘Measuring tax gaps’.

Hidden economy

Moonlighters

K.15 Moonlighters are defined as individuals who are employees in their legitimate occupation but do not declare earnings from other sources of income. There are two separate methodologies for different parts of the moonlighters estimate: one for earned income – that is individuals whose undeclared source of income is from employment – and one for unearned income – that is individuals whose undeclared source of income is not from employment but from sources such as lettings or interest.

K.16 To calculate the earned income estimate, data from the Hidden Economy Quantitative Survey (HEQS) was used. The survey was commissioned by HMRC in 2015 to understand the nature of the hidden economy and the characteristics of those involved. Data on prevalence and income from hidden economy activities was captured as part of this research. In total, 9,640 respondents were surveyed.
K.17 The estimate for unpaid tax on moonlighters’ earned income in the survey’s sample is calculated by subtracting the tax paid on declared income from the tax that would have been due on their earnings if they had declared all their income. This covers income tax and National Insurance Contributions (NICs), with allowances made for whether the hidden economy activity in question would be classified as self-employment or employment. An allowance for under-reporting of income is also made in line with academic literature.

K.18 This sample estimate is then grossed up to the total population by using the prevalence rates of moonlighters with earned income in the population. These prevalence rates are obtained from the HEQS and include weighting for non-response so that the prevalence rates are representative of the overall population.

K.19 A time series for the moonlighters’ earned income estimate was created by using a proxy index which took into account changes in receipts over time as well as data from the Family Resources Survey (FRS). The FRS is a government sponsored study which provides information about households in the UK. For more information go to www.gov.uk/government/collections/family-resources-survey--2.

K.20 The tax gap for moonlighters’ unearned income covers those individuals who have additional sources of income that are not from employment. These sources of income would therefore require them to submit a Self Assessment (SA) return to supplement their normal tax payment through PAYE.

K.21 The sources of income covered by unearned income are lettings, interest, capital gains on property, chargeable events, Individuals Savings Accounts (ISAs) and secondary income (for example, activities such as hobbies or online selling that are not regular enough to be considered employment).

K.22 It is not necessary for most taxpayers to submit a SA return where all tax liabilities are withheld at source. For example, employment income where tax is deducted under PAYE, or basic rate tax withheld from bank interest. However, there are risks within this population, for example due to taxpayers not informing HMRC about sources of income, especially where they may exceed tax-free allowances. Where a SA return should have been completed, lettings, interest and ISA income would be subject to income tax; capital gains on property and chargeable events would be subject to Capital Gains Tax (CGT); and secondary income would be subject to income tax and NICs.

K.23 HMRC cannot conduct random enquiries into the tax affairs of individuals who did not file a return because the legal position requires a return to be filed for an enquiry to take place. An alternative method is required for measurement of risks and estimating the associated tax gap.

K.24 HMRC has therefore used data matching of administrative data and third-party information to measure the extent to which taxpayers fail to declare these additional sources of unearned income, with an estimate of additional tax due being calculated from the identified undeclared income. Third-party data matched with administrative tax records includes rental deposit schemes and bank and building society interest declarations. Because of the large amount of data involved in this exercise, data matching is only conducted on a representative sample of the population already in PAYE. The results are thereafter grossed up from the sample to produce an estimate of the overall tax gap from moonlighters’ unearned income.

K.25 The limitations associated with the results of this exercise relate to the coverage of the third-party data used to establish evidence of additional undeclared income. Coverage varies across different sources of income, being especially good for lettings and interest income, whereas it is less reliable for the remaining sources identified. Additionally, there are other sources of income that could not be investigated due to unavailability of data. The resulting estimate should be interpreted broadly as a lower limit for the true scale of the tax gap relating to this group of taxpayers.

K.26 The latest estimate of the tax gap relating to moonlighters’ unearned income is for tax year 2014 to 2015. This is projected forward based on receipts changes over time taking into account policy
changes. For example, lettings income is subject to income tax; we take the lettings data-matching estimate for 2014 to 2015 and multiply it by a value which adds together income tax receipts for 2014 to 2015 and policy changes affecting receipts in 2015 to 2016 to obtain an estimate of how much policy changes have increased, or decreased the income tax receipts. This allows the projections to take into account changes in both tax rates and the tax base over time. For example, increases in the personal allowance reduce the potential tax revenue from hidden economy activities, all else being equal. The projections are based on the Office for Budget Responsibility’s certified costings estimates for all income tax, NICs and CGT policy measures, and the relevant tax regime is applied for each of the unearned income sources.

K.27 The 2017 to 2018 and 2018 to 2019 projected figures are further adjusted to take into account the new “tax free allowance for landlords” that was introduced in April 2017.

Ghosts

K.28 Ghosts are defined as individuals who do not declare any of their income to HMRC, be it earned or unearned.

K.29 Data from the HEQS is used to estimate the ghosts tax gap. See the moonlighters section beginning at paragraph K.16 for details.

K.30 The estimate for unpaid tax on ghosts’ income in the survey’s sample is calculated by applying the relevant tax rate to the undeclared income estimated from the survey observations. This covers income tax and NICs, with allowances made for whether the hidden economy activity in question would be classified as self-employment or employment. An allowance for under-reporting of income is also made in line with academic literature.

K.31 This sample estimate is then grossed up to the total population by using the prevalence rates of ghosts in the population. These prevalence rates are obtained from the HEQS and include weighting for non-response so that the prevalence rates are representative of the overall population.

K.32 As with moonlighters, a time series for the ghosts tax gap estimate was created by using a proxy index which took into account changes in receipts over time as well as data from the FRS.

K.33 Like moonlighters, 2017 to 2018 and 2018 to 2019 figures are adjusted to reflect the impact of new “tax free allowance on self-employed traders” that was introduced in April 2017.
Chapter L: Abbreviations

**BBPA:** British Beer and Pub Association

**BEIS:** Department for Business, Energy and Industrial Strategy

**CGT:** Capital Gains Tax

**CHIEF:** Customs Handling of Import and Export Freight

**CT:** Corporation Tax

**DOTAS:** Disclosure of Tax Avoidance Schemes

**EA:** Environment Agency

**EC:** Employer Compliance

**EU:** European Union

**EV:** Extreme values

**EX46:** Beer Duty Return

**FRS:** Family Resource Survey

**FSS:** Family Spending Survey

**GB:** Great Britain

**GLF:** General Lifestyle Survey

**HEQS:** Hidden Economy Quantitative Survey

**HMRC:** Her Majesty’s Revenue and Customs

**HRT:** Hand-rolling tobacco

**HSE:** Health Survey for England

**IhT:** Inheritance Tax

**IPS:** International Passenger Survey

**IRS:** Internal Revenue Service (United States)

**ISA:** Individuals Savings Account

**IT:** Income tax

**LCF:** Living Costs and Food Survey

**MSB:** Mid-sized business

**NAAFI:** Navy, Army and Air Force Institutes

**NHS:** National Health Service