

Office for
**Budget
Responsibility**

Working paper No.20

A review of welfare policy costings

Surjinder Johal

October 2023

A review of welfare policy costings

Surjinder Johal

Office for Budget Responsibility

Abstract

Since its creation, the OBR has scrutinised over 250 policy costings relating to changes in the benefits and tax credits systems. This paper looks back at how our scrutiny of those policy costings has evolved by reviewing ten case studies of major welfare policy changes taken from the 41 welfare policy announcements made by the Coalition Government in 2010 and included in our first two forecasts. We evaluate how these costings performed against our original expectations and explain where and why they deviated. In doing so we highlight some of the challenges we face when scrutinising welfare policy costings, and indeed costings more generally.

We also assess the aggregate fiscal impact of all the welfare policy announcements made in the June 2010 Budget and the November 2010 Spending Review and show that the expected reduction in welfare spending was around a third lower than we originally forecast.

We conclude the paper by highlighting the lessons we have learned over the intervening period and how those can be applied to improve our scrutiny of costings in the future.

We would like to thank Andy King, Laura Gardiner, Toby Irvine, Ellen Brett, Shaun Butcher, Samir Jahan, David Nolan, James Murphy Corkhill, the Budget Responsibility Committee and analysts in DWP and HMRC for their assistance and advice in preparing this paper.

Contents

Chapter 1	Introduction.....	1
	The OBR and the policy costings process	1
	The purpose of this paper	3
	Structure of the paper.....	4
Chapter 2	Welfare spending since 2010.....	6
	Introduction	6
	Welfare spending policy decisions since 2010	6
	Recosting the 2010 package of welfare measures.....	9
	The case studies.....	12
Chapter 3	Differences due to baseline forecast	18
	Introduction	18
	The economic determinants of welfare spending.....	19
	Case study 1: Measures subject to less generous uprating	20
	Case study 2: Housing benefit: increasing the age limit for the shared accommodation rate from 25 to 35	22
	Demographic factors affecting welfare spending.....	25
	Case study 3: Child tax credit: abolishing the baby element	26
	Other baseline factors affecting welfare spending	28
	Case study 4: Working tax credit: reducing support towards childcare costs from 80 per cent to 70 per cent.....	29
Chapter 4	Differences within the policy costing.....	33
	Introduction	33
	Data uncertainty	34
	Case study 5: Working tax credit: increase the working hours requirement for couples with children from 16 to 24 hours a week	35
	Modelling uncertainty.....	38

Case study 6: Contributory employment and support allowance: introduce a one-year time limit for those in the work-related activity group	41
Case study 7: Benefit cap	44
Direct behavioural effects in welfare costings	46
Case study 8: High income child benefit charge	48
Indirect behavioural effects and impacts on the labour market.....	51
Case study 9: Income support: extend lone parent conditionality to those with children aged five and above.....	54
Case study 10: Housing benefit: removal of the spare room subsidy	56
Chapter 5 Conclusions and lessons learned.....	59
Explaining deviations in welfare policy costings.....	59
How our scrutiny of welfare policy costings has evolved.....	61
Lessons for the scrutiny of future welfare costings.....	62

1 Introduction

The OBR and the policy costings process

- 1.1 The Office for Budget Responsibility (OBR) was created in 2010 to provide independent and authoritative analysis of the UK's public finances. To that end we produce two 5-year-ahead forecasts for the economy and the public finances each year, alongside Budgets and other fiscal statements, which we report in our *Economic and fiscal outlook (EFO)*.
- 1.2 Policy costings are an essential component of our forecast, and in each *EFO* we estimate the fiscal impacts of the Government's newly announced tax and spending policies. The Treasury presents an estimate of the effect of each measure on public sector net borrowing in the 'scorecard' it publishes alongside each statement. Under the *Charter for Budget Responsibility*, the Treasury is free to decide which measures to include in the scorecard and what costs or yields to attribute to them. In practice, the costing of every policy that has an impact on tax or annually managed expenditure (AME) goes through a detailed process of scrutiny and discussion between the OBR, the Treasury and the department responsible for implementing the policy (primarily HM Revenue and Customs (HMRC) for tax changes and the Department for Work and Pensions (DWP) for welfare measures).¹
- 1.3 The Treasury is required to provide the OBR with a policy costing note for every new tax and AME measure, containing a description of the policy, the methodology underpinning the costing, and a discussion of the uncertainties. Our role is to scrutinise the costing and decide whether we consider it to be a 'reasonable and central' estimate, a process that can take several rounds of scrutiny, with departments required to resubmit notes that respond to OBR queries and comments. Once we deem a costing is reasonable and central it is given a formal certification, which we confirm in our *EFO*. We then incorporate each costing (or, if we disagreed with the Treasury's published scorecard, our preferred ones – something we have not yet found necessary) in our forecasts, together with the impact of any relevant policy measures that the Treasury might have omitted from the scorecard.²
- 1.4 A policy costing note typically contains: an estimate of the tax or cost base to which the policy change is to apply; a 'static' costing (the fiscal impact in the absence of any resulting change in behaviour); plus the impact of any *direct* or 'first round' behavioural effects. For example, for a measure raising the income threshold for an existing benefit: the cost base is the current level of welfare spending on the affected population – those whose income now

¹ Our forecasts for spending within departmental expenditure limits ('DELS') typically follow the Government's announced levels of spending but with an allowance for shortfall. While we routinely highlight the risks around the likelihood of future revisions to those levels, we are required by Parliament to condition our forecasts on stated government policy. As we note in Atkins, G. and L. Lanskey, *The OBR's forecasting performance*, Working Paper No. 19, August 2023, underestimating the medium-term level of spending within DELs is one of the main factors contributing to our tendency to underestimate borrowing in the medium term.

² More detailed information on the costings process and methodology is presented in our *Briefing Paper No.6: Policy costings and our forecast*, March 2014.

takes them over the eligibility threshold; the static costing is the immediate reduction in welfare spending for that group; while a post-behavioural effect might involve some of those affected switching to an alternative benefit, for which they remain eligible.

- 1.5 In each forecast we also incorporate the *indirect* effects of the Government's policy decisions. These are 'second round' behavioural effects, those with wider-reaching, whole economy impacts. For demand side effects we use fiscal multipliers to estimate the combined effect on aggregate demand of all policy decisions in a fiscal statement. First-year multipliers vary for different categories of spending or tax, but all taper to zero by the end of our 5-year forecast period, in line with the empirical evidence. These demand-side effects are, therefore, assumed to be temporary.³
- 1.6 We also consider the economic impacts of policies individually, typically on the supply side of the economy. Supply-side effects, unlike those on the demand side, are assumed to be longer lasting. In the example above, some claimants might respond to the lowering or removal of their award by returning to the labour market or by increasing their hours worked. Such an effect can be incorporated either by a change in our labour market forecast (if it is deemed material) or through the costing itself. When evaluating the potential indirect effect of a specific policy, we apply defined criteria. Firstly, it must be backed by credible evidence, and the impact must be material for our forecast. We also consider the broader economic and policy landscape, ensuring that any impact is additional to what is captured by the outturn data (and therefore already included in our pre-measures forecast) and that the measure will have a lasting impact.⁴
- 1.7 To be transparent about the uncertainty of each costing, we assign each certified costing a subjective uncertainty rating.⁵ These range from 'low' to 'very high' and are based on our assessment of the uncertainty arising from each of three sources: the data underpinning the costing; the complexity of the modelling; and the possible behavioural responses to the policy change. We consider the relative importance of each source of uncertainty when determining the overall uncertainty rating. We report on the measures with the most uncertain costings in our *EFO*, explaining the reasoning behind our assessment and the risks to the forecast.
- 1.8 It is important to emphasise that, when we describe a certified costing as particularly uncertain, we see risks lying to both sides of what we nonetheless judge to be a reasonable and central estimate. Regardless of the level of uncertainty, we see a 50 per cent chance that the costing is higher than the central estimate, and a 50 per cent chance that it is lower. Others may differ in their views of the central estimate or the uncertainty around it, with the scope for different views likely to be greater for more uncertain costings.

³ We discuss fiscal multipliers in our December 2019 *Forecast evaluation report* and our November 2020 *Economic and fiscal outlook*.

⁴ For more, see *Briefing Paper No.6: Policy costings and our forecast*, March 2014, and *Briefing Paper No.8: Forecasting potential output – the supply side of the economy*, November 2022.

⁵ The complete set of uncertainty ratings can be found in our online *Policy costings uncertainty ratings database*.

The purpose of this paper

- 1.9 The OBR's remit includes a requirement to assess the performance of its forecasts. We do this annually in our *Forecast evaluation report (FER)*, where we compare the latest outturn data for the economy and public finances to our earlier forecasts, diagnosing and then explaining the differences. This process is invaluable in identifying lessons that we can take on board for future forecasts.
- 1.10 The same rationale applies with policy costings. The lessons we learn from reviewing previous costings informs future judgements on similar measures, helping to improve the quality of those costings. Some of those lessons may arise from formal reviews and evaluations but they may equally come from our forecast-to-forecast monitoring of costings and the recosting of previously announced measures that we carry out in every *EFO*. Our approach to policy costings has therefore evolved alongside our accumulated experience and our familiarity with the underlying evidence.
- 1.11 We have touched on some of the challenges associated with scrutinising welfare policy costings in previous editions of our *Welfare trends report (WTR)*. One of those is the difficulty in assessing delivery risk, especially from policies that require departments to implement operational changes over several years. More rigorous scrutiny of operational delivery is one of the key lessons we learned from our formative years, one that is mirrored from tax measures.⁶ We have retrospectively acknowledged that some costings, such as that for the 2010 reform to disability living allowance that led to the creation of personal independence payment, would not be certified now, unless accompanied by detailed evidence on how the stated policy objective would be achieved.⁷
- 1.12 In this paper we have selected ten welfare policy costing case studies from 2010 that allow us to investigate and illustrate some of the challenges involved in scrutinising the effect of policy changes in the benefits and tax credits systems, and how our approach has evolved. They were selected based on the range of lessons we can draw from analysing the differences between the original and revised costings and each also had a material impact on welfare spending. Nine of them were expected to reduce caseloads, by an uncertain amount, and involved changes in eligibility, or substantial changes in the benefit awards for those affected. The remaining case study (which combines six measures) has the largest fiscal impact but involves a simpler and more general policy change – to the annual rate by which benefit awards are uprated. It, and several others, show how factors within the baseline economic and fiscal forecast can cause a costing to differ significantly from its original estimate, even when the arithmetic of the costing itself is straightforward.
- 1.13 This is not a formal evaluation in the sense that we are not attempting to generate a full pre-measure counterfactual against which to re-cost the original measures. Instead, we have chosen a more general approach to estimating and analysing differences from

⁶ See Johal, S. and J. Sousa, *Anti-avoidance costings: an evaluation*, OBR Working Paper No. 8, January 2016 and Johal, S. *Evaluation of HMRC anti-avoidance and operational measures*, OBR Working Paper No.11, September 2017.

⁷ *Welfare trends report*, January 2019.

original costings, similar to the approach we use when recosting previously announced measures for our *EFO* forecasts. This allows us to demonstrate how and why estimates can deviate from those that we deemed to be reasonable and central at the time.⁸

- 1.14 We have used several different approaches to arrive at our revised estimates, including updated information provided to us by DWP and HMRC as well as additional analysis of our own. We are able to make use of the outturn data published by DWP, HMRC and the ONS. It is worth stressing that the outturn data, even when it maps across directly to a policy intervention (it rarely does) does not necessarily provide us with a conclusive answer as to the what the *actual* costing is, because the costing is measured against a *pre-measure* forecast of spending. Furthermore, when a costing has several components, making a definitive ex-post determination of how each has affected the difference between the original and revised estimate can be challenging. Therefore, each of our re-estimations carries a degree of uncertainty, with these uncertainties magnified for the decompositions that we carry out for the ten case studies (which explains why we include a residual).
- 1.15 The main lessons we draw from the paper are that: governments have a tendency to relax their initially tighter spending plans, leading to higher-than-expected welfare spending; we need to ensure that departments are able to operationally deliver a policy as it is laid out in the costing note, and pay as much attention to a measure's implementation as we do to its costing arithmetic; while the data and models that are used for welfare costings are generally reliable, the delay in data availability can be challenging in some instances, and that departments need to design appropriate administrative data collection and survey processes from the outset; and finally, departments (and the OBR) should conduct more costings evaluations to improve the evidence base for costings, which goes alongside the previous point around ensuring the right data is collected. It is also notable how surprises in the baseline economic and fiscal forecasts can cause a costing to significantly deviate from expectations even when each of the parameters in the costing itself performs as anticipated.

Structure of the paper

- 1.16 The rest of the paper is structured as follows:
- **Chapter 2** presents the aggregate impact of the 259 welfare policy measures announced since 2010, including revisiting the aggregate fiscal impact of the 41 measures that were announced by the Coalition Government in the June 2010 Budget and the November 2010 Spending Review. The Chapter also introduces the ten case studies from 2010 that we review in detail in Chapters 3 and 4.
 - **Chapter 3** explains how the costings of individual policies can deviate from their original estimates by unforeseen differences in the baseline economic and fiscal forecasts that underpin them. The first four of the case studies are used to illustrate this.

⁸ Reflecting the remit given to us by Parliament, we focus only on the impact of these policies on the public finances, and do not consider their distributional impact, the efficiency of their delivery, their value for money or their contribution to Governments' policy objectives.

- **Chapter 4** delves into the components of the policy costing itself, including uncertainties relating to the data, the modelling and any behavioural responses. It also discusses how and when welfare policy decisions might affect the wider economy. We support our analysis with the remaining six case studies.
- **Chapter 5** concludes and discusses some of the lessons we have learned since 2010 from scrutinising the costings of welfare policy changes.

2 Welfare spending since 2010

Introduction

2.1 In 2010 the Coalition Government introduced a package of policy measures designed to reduce the deficit, which in that year stood at 11 per cent of GDP. The bulk of the consolidation came from measures announced within the first six months of coming into office, at the June 2010 Budget and the November 2010 Spending Review. Relative to the more modest deficit reduction plans inherited from the previous Labour Government, the largest additional takeaways came from lowering current spending on public services, an increase in the main rate of VAT, and significant reductions to welfare spending. It is these welfare policy measures that are the focus of this paper.

2.2 In this chapter we:

- present the aggregate and cumulative impact of **the 259 welfare spending policy decisions that have been made from June 2010 onwards**;
- re-cost **the 41 measures that were announced in the June 2010 Budget and Spending Review 2010**, showing that the savings were around a third lower than initial expectations; and
- introduce **the ten case studies** that we have chosen to explore in detail in this paper, using them to explain how and why costings can deviate from their original estimates.

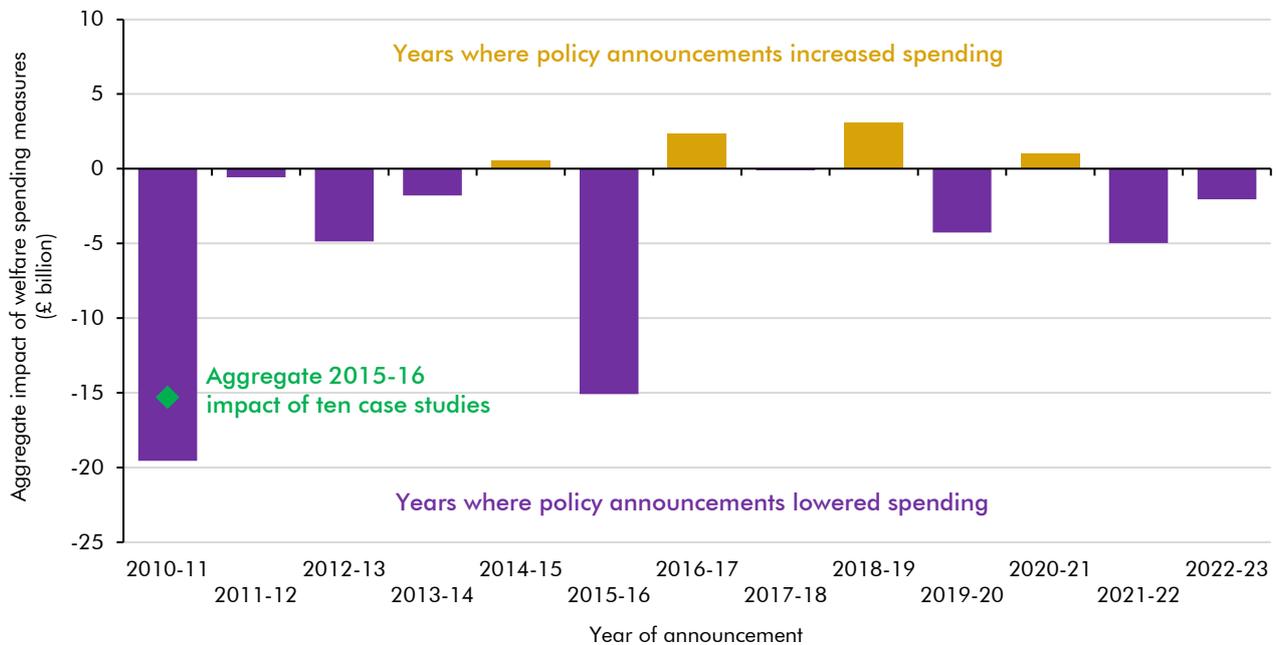
Welfare spending policy decisions since 2010

2.3 Chart 2.1 shows the aggregate effect on spending of all the welfare measures announced since June 2010, based on the final year of the original five-year scorecard.¹ As you can see, the largest reductions in spending came in the two Budgets immediately after the 2010 and 2015 general elections, with the measures announced in 2010-11 amounting to around £20 billion of savings in 2015-16, and those announced in 2015-16 lowering spending by around £15 billion in 2020-21 (£13.3 billion of which was announced in the July 2015 Budget). These amounts represented 8.1 and 6.2 per cent of the respective pre-measures welfare spending forecasts at the time. (Most of the savings related to working-

¹ For example, the first bar shows the 2015-16 impact of the measures announced in the three fiscal events during 2010-11 – the June 2010 Budget, November 2010 Spending Review and March 2011 Budget. The published scorecard for the June 2010 Budget only extended to 2014-15, but costings for 2015-16 were provided to us at the time to complete the five-year forecast, so they have been used in Charts 2.1 and 2.2. Also note that these charts only include the impact on welfare spending, rather than the overall scorecard impact, so excludes the modest knock-on implications of welfare measures for tax receipts. One limitation of presenting costings based on their original scorecard estimate is that it does not factor in temporary (or permanent) reversals to policies, or temporary boosts to welfare spending that offset the savings from the initial estimates. But these subsequent decisions are captured in later scorecards and, as we see in both Chart 2.1 and 2.2, the tendency since 2010 has been to announce policies that reduce overall welfare spending.

age and child welfare spending, with the reductions amounting to around 12.0 and 11.6 per cent of non-pensioner welfare spending, respectively.)

Chart 2.1: Welfare spending: final-year scorecard costings by year of announcement



2.4 The years with net giveaways include:

- **2016-17**, which was largely due to the decision not to implement a Budget 2016 measure that sought to lower spending on disability benefits via a reduction in the entitlement points that would be awarded in personal independence payment (PIP) for cases involving the use of certain aids and appliances;
- **2018-19**, mostly due to a partial policy reversal relating to the July 2015 package of welfare measures, delivered via a £1,000 increase in universal credit work allowances;² and
- **2020-21**, related to the small number of pandemic support measures that are still expected to have a fiscal impact in 2025-26. The cost of pandemic support measures (such as the additional £20 a week for recipients of universal credit) was much higher during 2020-21 and 2021-22, when they increased welfare spending by £8 billion and £5 billion, respectively. Relatively few of the support measures had a material fiscal impact beyond 2021-22, which is why the 2025-26 cost is so much lower.

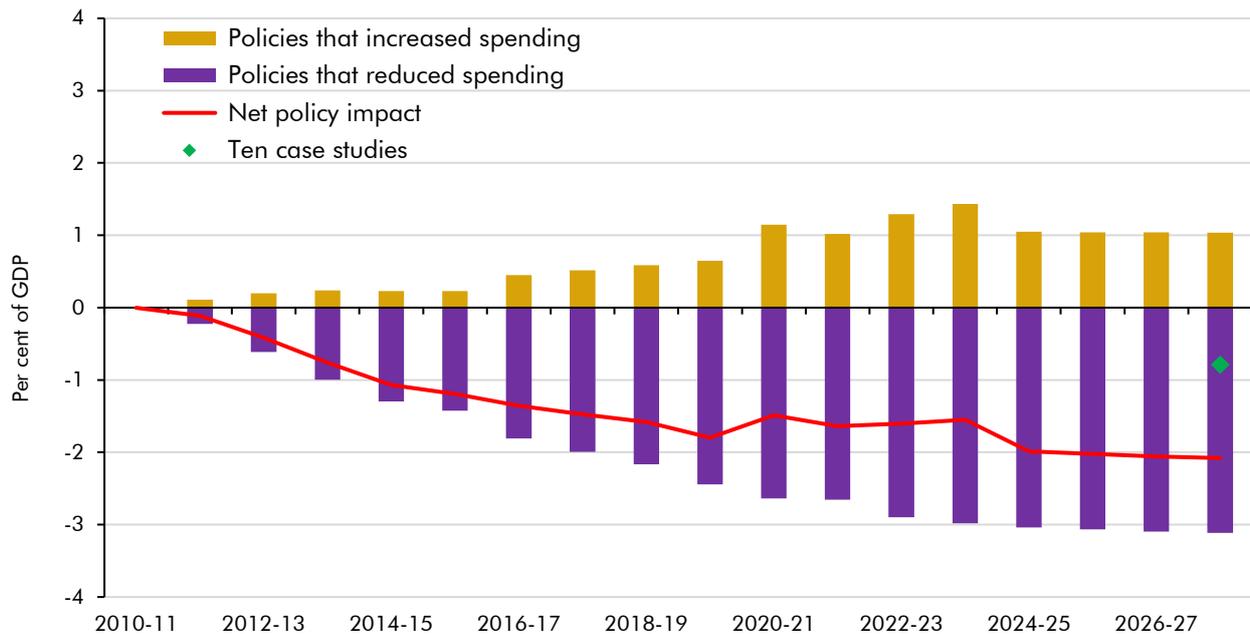
2.5 Chart 2.2 shows the cumulative effect on welfare spending of all policy announcements from June 2010 onwards, initially based on the original five-year scorecard costing and

² Welfare spending was also increased by the non-implementation of the October 2017 decision that funding for short-term supported accommodation would be provided via a grant, rather than continuing to be met through housing benefit and universal credit. The switch was fiscally neutral, with the reduction in welfare spending matched by an increase in spending within departmental budgets.

Welfare spending since 2010

then grown in line with nominal GDP thereafter.³ On this basis, the aggregate effect of all welfare measures announced since 2010 amounts to a net takeaway of 2.1 per cent of GDP by 2027-28, or 28 per cent of our current forecast for welfare spending in that year. Chart 2.2 also presents the aggregate and cumulative savings from our 10 case studies, produced on the same basis, which peak at 0.8 per cent of GDP.

Chart 2.2: Welfare spending: cumulative impact of measures since June 2010

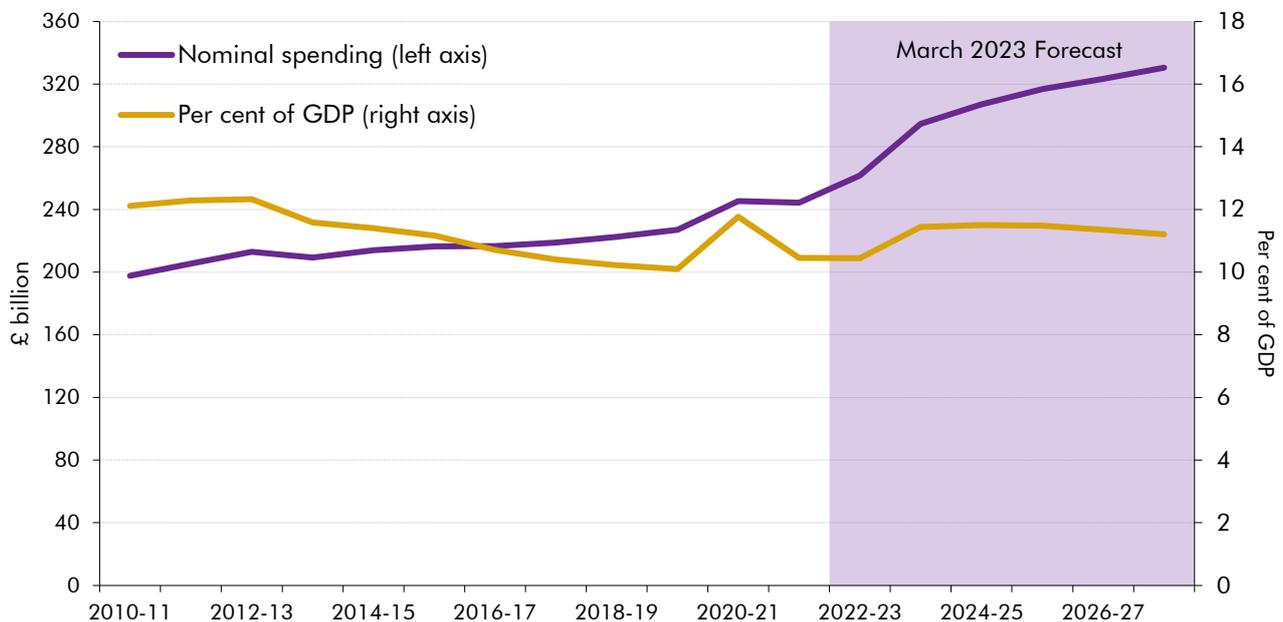


Source: OBR

2.6 For context, Chart 2.3 shows the evolution of overall welfare spending over the same time period. On a nominal basis (so like-for-like with Chart 2.1) welfare spending has continued to rise, though it has fallen slightly when we look at it as a percentage of GDP (the measure in Chart 2.2). But this fall is less than that implied by the accumulation of individual scorecard measures above, and as we explain below, both the 2010 and 2015 packages lowered welfare spending by considerably less than expected.

³ This approach is consistent with our online *Policy measures database*. While this is a reasonable approximation for the majority of costings, it performs less well for changes to the uprating of statutory benefit rates, where the effect continues to grow substantially due to systematic differences between uprating indices, and for policies that are not fully mature at the end of the forecast horizon. This approach does, however, allow for the separate inclusion of subsequent policy reversals in future scorecards.

Chart 2.3: Total welfare spending, 2010-11 to 2027-28



Source: ONS, OBR

2.7 Charts 2.1 and 2.2 add up the value of the original scorecard costings at the time of each policy announcement, which serves as a useful and practical benchmark against which to measure our revised estimates. This approach does not, therefore, take account of whether those initial costings were accurate and, as we stress in every *Economic and fiscal outlook*, there is uncertainty around the central estimate of the impact of each policy measure. But in each forecast we do include the recostings of previously announced measures, so any divergence from the original costing is usually captured relatively quickly.

Recosting the 2010 package of welfare measures

2.8 There were 28 welfare measures announced in the June 2010 Budget, with a further 13 announced in November 2010's Spending Review. Tables 2.1 and 2.2 list the different measures, along with the original and revised estimates of the costings in 2014-15, which was the final year of the June 2010 published scorecard and the year we focus on in this paper.⁴ The tables also highlight the 15 scorecard measures that form our 10 case studies, which have been selected based on the lessons we can draw from the differences between the original and revised costings for our welfare costings process.

2.9 We have used several different approaches to arrive at our estimates for these case studies, including recostings (some of which were published in Budget documents after 2010) or updated information provided to us by departments, as well as additional analysis of our own. For some smaller measures that were implemented during 2010-11, for simplicity we use the original scorecard costing, since that remains our best estimate.

⁴ Tables 2.1 and 2.2 present the overall scorecard impact of each measure. While welfare spending makes up the vast majority of the fiscal impact of these measures, there are other, smaller impacts, such as on tax receipts.

2.10 Table 2.1 shows that the June 2010 package consisted of 23 measures that sought to reduce welfare spending and five that were expected to increase it. The net aggregate 2014-15 scorecard impact of the original costings was a saving of £10.9 billion (£7.2 billion from those within our case studies). This is made up of £13.5 billion of total savings from those measures that reduce spending (of which £8.2 billion is from our case studies) and a total cost of £2.6 billion from the measures that increase spending (£1.0 billion from the case studies).

2.11 Table 2.2 shows the equivalent information for the November 2010 package.⁵ The net aggregate scorecard impact of those measures in 2014-15 was expected to be a saving of £7.5 billion (an £8.3 billion gross saving from measures reducing spending, partly offset by a £0.8 billion cost from those raising spending). The case studies from November 2010 are all measures that sought to reduce welfare spending, by an expected £6.8 billion in 2014-15.⁶

2.12 Our latest assessment is that:

- **The June 2010 measures reduced spending by £9.0 billion in 2014-15, £1.9 billion (18 per cent) fewer savings than originally estimated**, with around £0.8 billion (around two-fifths) of the difference coming from the case studies. The remaining shortfall is largely due to the measure that introduced personal independence payment as the replacement for disability living allowance, which fell far short of the Government's aim to reduce spending materially. That shortfall was explored in detail in our January 2019 *Welfare trends report (WTR)* and is discussed briefly in paragraph 2.15 below.
- **The November 2010 measures reduced spending by £3.8 billion in 2014-15, but this is £3.6 billion (49 per cent) less than initially expected**, almost all of which relates to shortfalls in savings expected from the case study measures.
- **The combined fiscal impact across all the measures announced at these two post-election events in 2010 is £12.8 billion, which means that borrowing was £5.6 billion (30 per cent) higher than we assumed** in our 2010 forecasts.

2.13 This result echoes that reported in our December 2019 *WTR*, which assessed the costings of the July 2015 Summer Budget package of welfare measures. We showed that those measures saved around a third less than expected, with a revised estimate of £9.1 billion, down from the original £13.3 billion. We concluded at the time that this was due to *“various policy reversals, delays and concessions related to cuts to tax credits and [universal credit]”*, with the reversed policies *“typically those that generated cash losers from one year to the next”*.

⁵ One welfare measure from 2010, relating to support for mortgage interest, was not expected to have a fiscal impact in 2014-15 so has been omitted.

⁶ For some measures, the original estimate in Table 2.1 and 2.2 is different to that presented in their respective case studies. The reason is that the latter includes the impact of amendments to the original policies. The rationale for the different treatments is that Tables 2.1 and 2.2 are looking into how much the overall savings in net welfare spending fell below initial expectations, and one of the reasons that policies sometimes bring in less than expected is that they are either watered down by subsequent revisions or reversed completely. On that basis it is reasonable to measure the revised costing directly against the original. By contrast, the case studies are exploring the analytical (or non-policy) reasons that costings can deviate from their original estimates, and for that it is sensible to account for policy amendments by first restating the original costing before generating our diagnostics. In that way we are better able to evaluate the true fiscal impacts of each of the analytical components within the costing and the baseline forecast that underpins it.

Table 2.1: June 2010 Budget: original and revised costings for 2014-15

	£ million		
	Original	Revised	Difference
Benefits, tax credits and public service pensions: switch to CPI indexation from 2011-12	-5,845	-3,880	1,965
Child tax credit: increase the child element by £150 in 2011-12 and £60 in 2012-13 above indexation	1,595	1,595	0
Disability living allowance: reform gateway from 2013-14	-1,075	-40	1,035
Child benefit: freeze for 3 years from 2011-12	-975	-1,375	-400
First and second withdrawal rates: increase to 41% in 2011-12	-765	-755	10
Introduce an income disregard of £2,500 for falls in income from 2012-13	-585	-475	110
Pension credit minimum income guarantee: matching basic State Pension cash increase in 2011-12	535	650	115
Social sector: limit working age entitlements to reflect size of family from 2013-14*	-490	-365	125
Child tax credit: taper the family element immediately after the child element from 2012-13	-480	-425	55
Basic State Pension: introduce triple guarantee from 2011-12	450	-645	-1,095
Local housing allowance: set at the 30th percentile of local rents from 2011-12	-425	-475	-50
Reduce the income disregard from £25,000 to £10,000 for 2 years from 2011-12 then to £5,000 in 2013-14	-420	-540	-120
Switch to CPI indexation for local housing allowance from 2013-14	-390	-270	120
Deductions for non-dependents: reverse previous freezes and maintaining link with prices from 2011-12	-340	-210	130
New claims and changes of circumstances: Reduce backdating from 3 months to 1 month from 2012-13	-330	-360	-30
Child tax credit: remove the baby element from 2011-12	-275	-230	45
Lone parent benefits: extend conditionality to those with children aged 5 and above from October 2011	-180	-275	-95
Child tax credit: reverse the supplement for children aged one and two from 2012-13	-180	-180	0
Health in pregnancy grant: abolish	-150	-150	0
Tax credits second income threshold: reduce to £40,000 in 2011-12	-145	-135	10
Council tax: reduction to receipts due to a one year freeze in 2011-12	-135	-135	0
Reduce awards to 90% after 12 months for claimants of Jobseekers Allowance	-110	-110	0
Sure start maternity grant: apply to first child only from 2011-12	-75	-75	0
Local housing allowance: caps on maximum rates for each property size, with 4-bed limit from 2011-12	-65	-65	0
Support for mortgage interest: set payments at the average mortgage rate from October 2010	-65	-65	0
Working tax credit: remove the 50 plus element from 2012-13	-40	-50	-10
Additional discretionary housing payments	40	40	0
Additional bedroom for carers from 2011-12	15	15	0
TOTAL	-10,905	-8,985	1,920
Case studies	-7,170	-6,390	780

* The costing of this measure is consistent with the June 2010 Budget scorecard but is different to that presented in Case Study 10, where the 'original' estimate includes the impact of an amendment to the policy at Budget 2012.

Note: This table is in borrowing convention, i.e. a negative sign means a lowering in borrowing in 2014-15. Case studies are in bold.

Table 2.2: 2010 Spending Review: original and revised costings for 2014-15

	£ million		
	Original	Revised	Difference
Child benefit: remove from families with a higher rate taxpayer from January 2013	-2,500	-1,090	1,410
Contributory employment and support allowance: time limit for those in the work related activity group to one year	-1,970	-445	1,525
Working tax credit: freeze in the basic and 30 hour elements for three years from 2011-12	-835	-1,170	-335
Child tax credit: increase the child element by £30 in 2011 and £50 in 2012	765	690	-75
Working tax credit: increase working hours requirement for couples with children to 24 hours	-535	-115	420
Council tax benefit: 10% reduction in expenditure and localisation	-520	-520	0
Working tax credit: reduce payable costs through childcare element from 80% to 70% restoring 2006 rate	-500	-240	260
Savings credit: freeze maximum award for four years from 2011-12	-395	-415	-20
Child and working tax credits: use real time information	-385	-200	185
Total household benefit payments capped on the basis of average take-home pay for working households	-270	-80	190
Housing benefit: increase age limit for shared room rate from 25 to 35	-215	-140	75
Disability living allowance: remove mobility components for claimants in residential care	-160	-160	0
Cold weather payments: increase rate permanently to £25 from November 2010	50	50	0
TOTAL	-7,470	-3,835	3,635
Case studies	-6,825	-3,280	3,545

Notes:

1. The costings presented here are generally from our November 2010 *Economic and fiscal outlook*, which updated the scorecard that was published in Spending Review 2010, in line with our latest forecast at the time. The two exceptions are the measures relating to contributory ESA and the childcare element of working tax credit (to match the treatment in their respective case studies).
2. The costings for the removal of child benefit for higher rate taxpayers and the benefit cap are consistent with their original 2010 estimates, but that makes them different to the 'original' estimate as presented in their case studies. The difference is because the latter includes the impact of policy decisions subsequent to the initial announcement but prior to 2014-15.
3. The accounting treatment in respect of reduced liability tax credits has since changed.
4. The costing of the council tax benefit measure was assumed to be neutral in spending terms, since it abolished the benefit and replaced it with support from local authorities - the net impact reported here is therefore the tax impact. We have not revisited the amount of equivalent support provided by local authorities, implicitly assuming the change was indeed neutral.
5. This table is in borrowing convention, i.e. a negative sign means a lowering in borrowing in 2014-15.
6. Case studies are in bold.

The case studies

2.14 In Chapters 3 and 4 we carry out a detailed assessment of 10 case studies to explore the reasons why the actual savings from the 2010 package of measures were less than expected. Each of the 15 measures that make up the 10 case studies were fiscally significant, with an impact of at least £100 million in at least one scorecard year. They were also chosen because they help to illustrate some of the challenges we face when scrutinising costings more generally, and some of the lessons we have learned over the 13 years that the OBR has been performing this task.

2.15 The 10 case studies can broadly be grouped into two groups: (i) those where the differences between the original and revised costings can be explained by exogenous changes in the baseline economic and fiscal forecast (such as to inflation or earnings); and (ii) those aspects that are endogenous to the costing itself (such as the scale of the expected behavioural response). Nine of the case studies were expected to reduce caseloads, either directly or indirectly, and involved changes in eligibility, or substantial changes in the benefit awards for those affected. The remaining case study (which combines six measures) has a larger fiscal impact but involves a simpler and more general policy change – to the annual rate by which benefit awards are updated.

2.16 In choosing our case studies we have excluded the two major reforms to the welfare system that were announced in 2010, which are:⁷

- **The introduction of universal credit (UC)**, whose forecast we reviewed in depth in our 2018 *WTR*. That analysis fell short of a full evaluation, partly because it is difficult to disentangle the fiscal effects of UC from the many other welfare measures announced over the same period, including those within the legacy benefits and tax credits systems, as well as changes to the design of UC itself. One of the costings lessons we have learnt since 2010 is that complicated reforms of this kind are susceptible to significant implementation delays, which can bring with them additional costs. The current UC rollout end-date is beyond our latest forecast horizon, so no earlier than 2028, which puts it at least 11 years behind schedule.
- **The replacing of the disability living allowance with a new personal independence payment**, which we covered in our January 2019 *WTR*. In that report we showed that PIP is estimated to cost somewhere between £1 billion and £2 billion a year *more* than the system it replaced, even though it was expected to *reduce* spending by around £1.5 billion a year. As with UC, the rollout of PIP has not gone according to plan and is also now due to finish beyond our most recent medium-term forecast horizon (i.e. after March 2028). At a minimum, that puts the full PIP rollout 12 years behind schedule. Tables 2.1 and 2.2 show that our 10 case studies and this measure combined explain 96 per cent of the absolute difference between the original and revised costings in 2014-15.⁸

Case studies affected by differences in the baseline forecast

2.17 There are four case studies for which the difference between the original and latest costings is largely explained by differences between forecast and outturn in the underlying economic and fiscal forecasts that determine the baseline level of welfare spending. These four case studies, which are analysed in Chapter 3, are:

⁷ While the delays to UC and PIP are largely due to factors relating to implementation, more recently they do include discretionary policy decisions to slow the pace of the respective rollouts (decisions that typically reduced medium-term welfare spending).

⁸ UC does not factor into our re-estimation since, while it was announced in 2010, the announcement was not accompanied by a scorecard costing, with the Government saying in Spending Review 2010 that the “policy is at an early stage of development and key policy parameters are yet to be finalised”.

- 1 **Measures introducing less generous uprating of benefits.** This case study includes five measures from June 2010 and a further one from November 2010. In most cases the measures involved generating savings by lowering the uprating assumption, either by switching from RPI inflation to (the typically lower) CPI inflation or by imposing a freeze. They have been grouped into a single case study because the respective differences between the original and revised costings relate to surprises relative to our inflation forecast. These six measures were expected to generate a net saving of £7.1 billion in 2014-15 (£8.0 billion of gross savings offset by a close to £1 billion cost for the two triple lock measures). Our latest estimate puts the net saving at £6.7 billion, 5 per cent lower than expected (see case study 1 in Chapter 3).
 - **The uprating of most benefits was switched from RPI to CPI indexation** following the June 2010 decision to lower the default uprating assumptions for benefits, tax credits, housing benefit eligible rents, and public service pensions from 2011-12 onwards.⁹
 - Two measures imposed a three-year freeze from 2011-12 instead of uprating by CPI: for **child benefit** (announced in June 2010) and the **basic and 30-hour elements of working tax credit** (from November 2010).
 - June 2010's **basic state pension triple lock** changed the default uprating assumption from RPI to the higher of CPI inflation, earnings growth or 2.5 per cent, from April 2011 (though RPI indexation was used for 2011-12).¹⁰ Alongside it came **the pension credit minimum income guarantee**, which matched the basic state pension cash increase in 2011-12, ensuring that lower-income pensioners benefited from the triple lock by raising the pension credit standard income guarantee by a larger percentage than the default (earnings) indexation in April 2011 would have delivered.¹¹
 - **The local housing allowance (LHA) switch from RPI to CPI indexation** was one of three changes to LHA rates announced in June 2010.¹² From April 2013 rates were to be uprated annually at the lower of CPI and the 30th percentile of local market rents, with caps on the maximum rates for each property size, and a four-bed limit. Prior to this, rates were set on the basis of the median market rent and with a five-bed limit.
- 2 **Housing benefit: increasing the age limit for the shared accommodation rate (SAR) from 25 to 35.** This measure, announced at the 2010 Spending Review and effective

⁹ The six original scorecard names are: 'benefits, tax credits, and public service pensions: switch to CPI indexation from 2011-12'; 'child benefit: freeze for 3 years from 2011-12'; 'working tax credit: freeze in the basic and 30 hour elements for three years from 2011-12'; 'basic State Pension: introduce triple guarantee from 2011-12'; 'pension credit minimum income guarantee: matching basic State Pension cash increase in 2011-12'; and 'lone parent benefits: extend conditionality to those with children aged 5 and above from October 2011'.

¹⁰ This replaced a 'double lock' policy whereby the State Pension was increased by the higher of Rossi inflation and 2.5 per cent. Rossi inflation is RPI inflation minus housing costs (such as rents, mortgage interest payments and council tax).

¹¹ Similar measures were announced in Autumn Statement 2011, Budget 2013 and Autumn Statement 2013, though in those cases the threshold of the pension credit saving credit was increased by more than the default indexation so that the fiscal effects were broadly neutral. Our case study only considers the impact of the initial 2010 costing.

¹² Only the CPI indexation measure is included in the case study, though the recostings of the other two are shown in Table 2.1.

from April 2012, expanded the number of single people subject to the SAR (referred to at the time as the shared room rate). The SAR restricts the maximum housing benefit entitlement for single people living in the private-rented sector and claiming housing benefit under LHA rules. These claimants could now only receive the (lower) rate for a single room in a shared house, rather than the rate for a self-contained single bedroom property, a loss of £40 a week at the time. The original costing estimated savings of £215 million in 2014-15, which we have revised down to £140 million, 35 per cent lower (see case study 2 in Chapter 3).

- 3 **Child tax credit: abolishing the baby element.** This June 2010 Budget measure removed the additional payment under the child tax credits system for those with a child aged under 1, with effect from April 2011. This amounted to an annual loss of £545 per child of that age. The measure was originally estimated to save £275 million in 2014-15; our revised estimate is 16 per cent lower at £230 million (see case study 3 in Chapter 3).
- 4 **Working tax credit: reducing support towards childcare costs from 80 per cent to 70 per cent.** Households eligible for working tax credit could receive a payment covering a percentage of their childcare costs up to a maximum of £175 a week for households with one child and £300 a week for those with two or more children. This 2010 Spending Review measure lowered the proportion of costs that could be reclaimed from 80 to 70 per cent from April 2011, and was originally expected to save around £500 million in 2014-15.¹³ Our revised estimate for 2014-15 on the current accounting basis is £240 million, just under half the original estimate (see case study 4 in Chapter 3).

Case studies affected by factors in the policy costing

2.18 The remaining six case studies, which we discuss in detail in Chapter 4, are those where the differences from the original costings derive from the parameters within the costing itself, such as take-up rates and other behavioural responses to policy changes. The six are:

- 1 **Working tax credit: increasing the working hours requirement for couples with children from 16 to 24 hours** (as well as requiring one partner to work a minimum of 16 hours). The original costing expected savings to reach £535 million in 2014-15. Our latest estimate is that the measure reduced spending by 79 per cent less than originally expected at £115 million (see case study 5 in Chapter 4).
- 2 **Contributory employment and support allowance (ESA): introducing a one-year time limit for those in the work-related activity group**, from April 2012. The original costing anticipated savings rising to £2.0 billion in 2014-15. Our latest estimate for that year suggests a saving of just £0.4 billion, a 77 per cent shortfall, the largest of any of the measures considered in this paper (see case study 6 in Chapter 4).

¹³ This was made up of £385 million of lower spending and £115 million of lower negative tax. Under the current accounting treatment, this would all be reflected in lower spending, which is how we consider it here and is the basis for the revised costing. The Treasury's scorecard in the 2010 Spending Review only showed the effect on spending, while our subsequent November 2010 forecast captured the full impact.

- 3 **The 'benefit cap' limiting the total amount that a single household can receive from certain benefits** from April 2013.¹⁴ It applied to households containing working-age benefit recipients who are either out of work or ineligible for working tax credit. The original costing expected savings to rise to £265 million in 2014-15. Our latest estimate is 70 per cent lower at £80 million (see case study 7 in Chapter 4).
- 4 **The high-income child benefit charge introducing means-testing to child benefit**, withdrawing eligibility from higher-earning households. It was formally introduced at the March 2012 Budget, amending an earlier decision from the 2010 Spending Review.¹⁵ The original 2010 measure was expected to reduce spending by £2.5 billion in 2014-15, though the 2012 amendment lowered the estimate to £1.7 billion. Our revised costing of the latter suggests the actual savings have been almost 40 per cent lower at £1.1 billion, some of which has been realised through increased tax receipts rather than reduced spending (see case study 8 in Chapter 4).
- 5 **Income support: extend lone parent conditionality to those with children aged five and above.** This measure reduced the number of single parents that were eligible for income support from October 2011. Those single parents that were deemed able to work, and whose youngest child was aged five or six, were to be transferred to jobseeker's allowance. Those that were unable to work would instead transfer to ESA or be allowed to remain on income support. Savings in 2014-15 were originally expected to be £180 million, but our revised estimate is 53 per cent higher, at £275 million (see case study 9 in Chapter 4).
- 6 **Housing benefit: removal of the spare room subsidy** (more commonly known as the 'bedroom tax').¹⁶ This measure reduced housing benefit payments for working-age tenants in the social-rented sector who were deemed to "occupy a larger property than their family size warrants". It was due to begin in April 2013 and originally expected to reduce spending by £490 million in 2014-15. Two policy amendments before it was introduced increased the saving to £550 million. We now estimate that it reduced spending by £365 million, around a third lower (see case study 10 in Chapter 4).

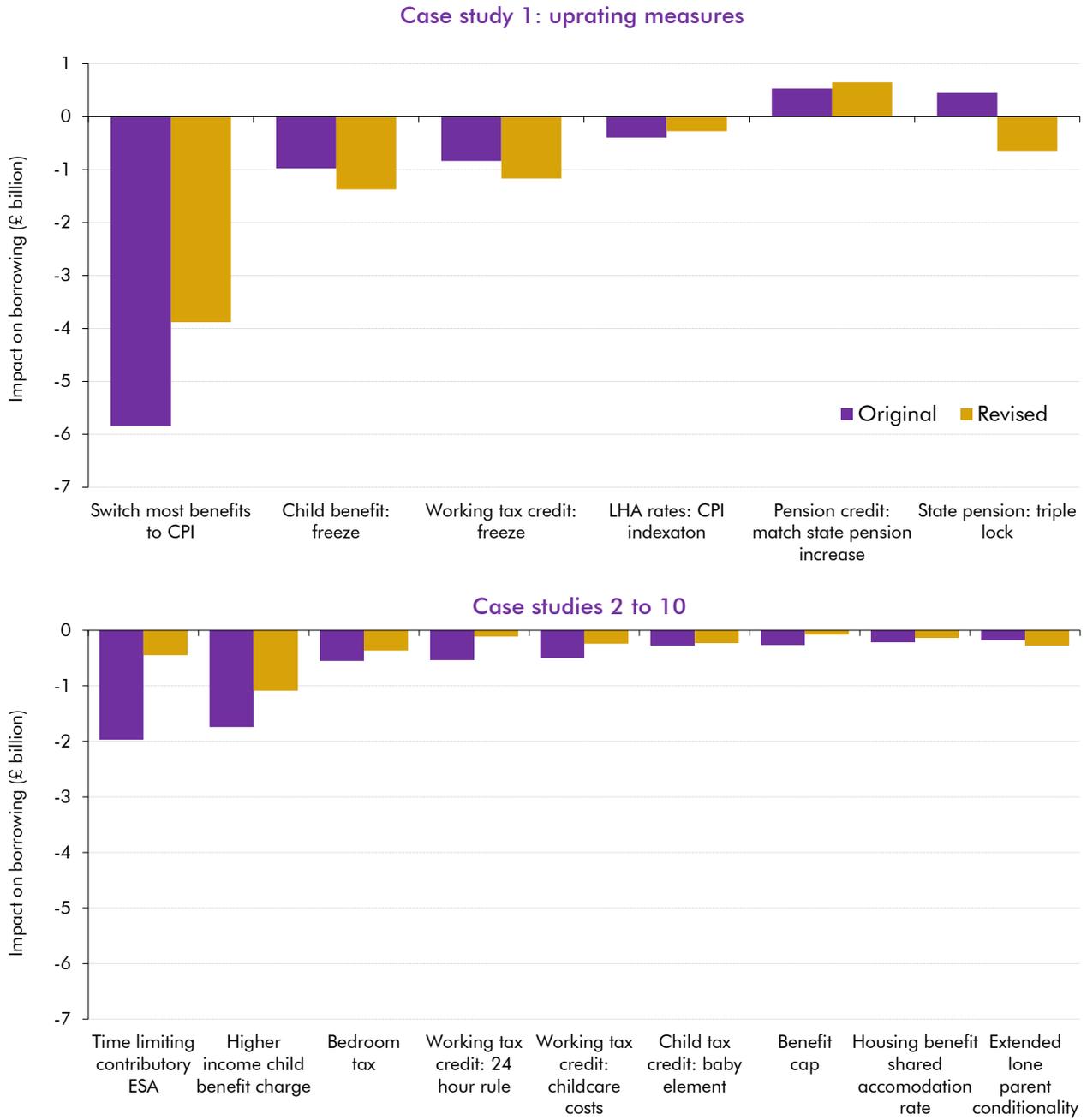
2.19 Chart 2.4 shows the original and revised 2014-15 costings for our 10 case studies. In aggregate, we have revised down the estimated savings from £14.0 billion to £9.7 billion, a shortfall of £4.3 billion (31 per cent). A single measure – the time-limiting of claims for contributory ESA – is responsible for around a third of the shortfall. Overall, the revised estimates are lower in nine of the 10 case studies and higher in just one.

¹⁴ Its original scorecard name was 'total household benefit payments capped on the basis of average take-home pay for working households'.

¹⁵ Its original scorecard name at Spending Review 2010 was 'child benefit: remove from families with a higher rate taxpayer from January 2013', while the Budget 2012 amendment was 'child benefit threshold and taper'. Table 2.2 contains the costing of the original measure.

¹⁶ Its original scorecard name was 'social sector: limit working age entitlements to reflect size of family from 2013-14'.

Chart 2.4: Original and revised 2014-15 costings of the 10 case studies



Source: OBR

3 Differences due to baseline forecast

Introduction

- 3.1 Our pre-measures economic forecast of the components of welfare spending provides the baseline for every welfare policy costing. Holding all else equal, any differences between this baseline economic forecast underpinning a costing and the eventual outturn will cause a costing to deviate from the original estimate. This means that even if all the components and judgements within the costing itself were borne out in the outturn data, so that the proportionate effect of the policy were correct, the effect in cash terms would still differ.
- 3.2 Our welfare spending forecast relies on several key economic determinants, including CPI inflation, average earnings and the rates of employment and unemployment, as well as non-economic factors, such as those relating to the operation of the benefits system and the composition of the population. The forecast of each of these is subject to uncertainty.¹
- 3.3 The baseline forecast also includes the costings of previously announced policies, including some yet to be fully implemented, which adds an additional layer of uncertainty.
- 3.4 In this chapter we:
- show how our 2010 forecasts of **the three main economic drivers of welfare spending** – inflation, earnings and employment – compare to the latest outturn data;
 - show how the ONS’ **population projections** that fed into our baseline forecasts in 2010 performed relative to the latest population estimates;
 - discuss some **key judgements** we take when forecasting baseline welfare spending, such as the eligibility for, and take-up of, different benefits, as well as factors unrelated to the benefits system, such as the number of people in full-time education; and
 - present **the recostings of four of the 10 case studies** described in Chapter 2, illustrating how these have been affected by differences from the underlying baseline forecasts.

¹ Our annual *Forecast evaluation report* details how our forecasts performed relative to the outturn data and the underlying reasons behind the differences. Our bi-ennial *Welfare trends report* provides more in-depth analysis specific to our welfare forecast. Our *Forecast revisions database* contains a decomposition of revisions to all our fiscal forecasts since 2010.

The economic determinants of welfare spending

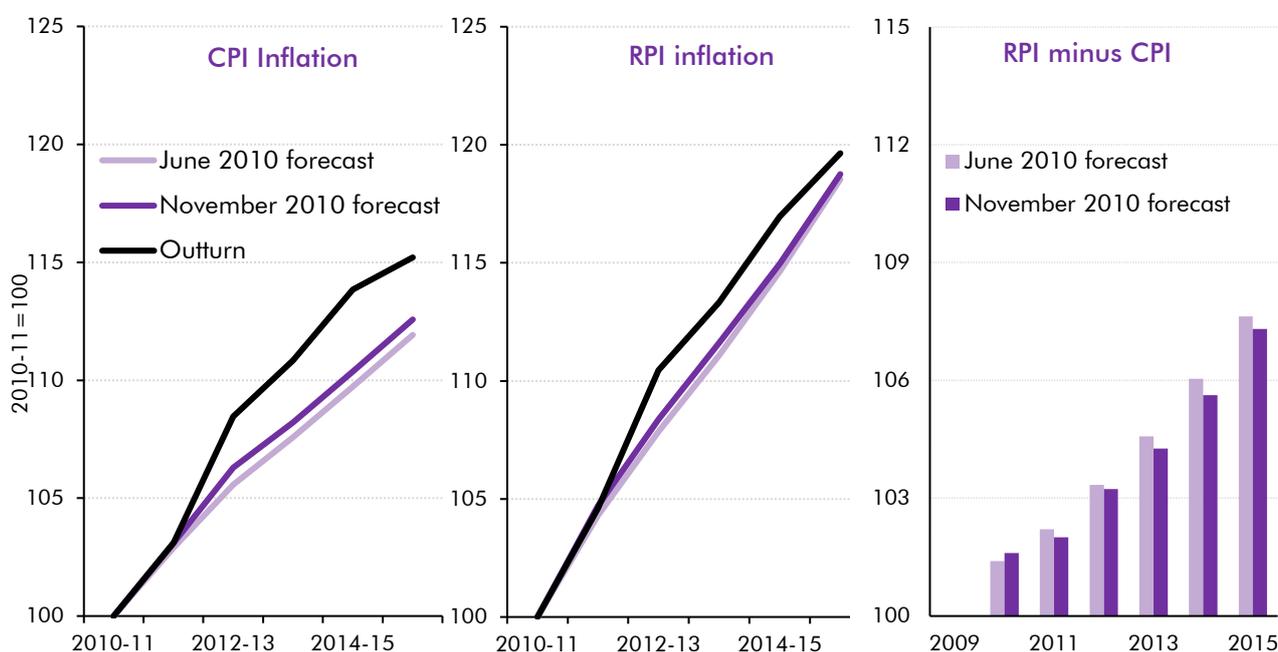
- 3.5 Our economy forecast is a key input into our fiscal forecasts, and developments in the economy explain much of the movement in the public finances from year to year. For example, in Annex B of our March 2016 *Economic and fiscal outlook (EFO)* we showed that revisions to our forecasts for nominal GDP growth had explained around 80 per cent of the variation in the revisions to our borrowing forecasts since June 2010.
- 3.6 The economy took a very different path to the one that we forecast in 2010, affecting all the case studies analysed in this paper.² For example, cumulative growth of real GDP in the five years to 2015-16 was 2.0 percentage points less than we expected in June 2010 (at 12.6 rather than 14.6 per cent). Over the same period, cumulative average earnings growth was 8.1 percentage points lower than forecast (10.8 versus 18.9 per cent), whereas cumulative employment growth was 3.5 percentage points higher (7.3 versus 3.8 per cent).

Inflation

- 3.7 The rates of most social security benefits and tax credits are uprated each year in line with CPI inflation, making it a key determinant of welfare spending. Prior to the policy changes announced in 2010, uprating was typically linked to RPI inflation, making that a key determinant in our baseline forecast at the time. Chart 3.1 compares our June and November 2010 forecasts for RPI and CPI inflation with the latest ONS outturn data.
- 3.8 CPI inflation was higher than we expected in 2010 across most of the forecast period. The cumulative increase was 15.2 per cent by 2015-16, which is 3.3 and 2.6 percentage points higher than our June and November 2010 forecasts, respectively. In part, this reflects the large movements in oil prices over the period. For example, the 5.2 per cent CPI inflation rate in September 2011, driven by higher oil prices, that was used for benefit uprating in 2012-13 is a key factor behind that year turning out to be the all-time peak in welfare spending as a share of GDP. Subsequent policy measures that either froze working-age benefits or limited their uprating to 1 per cent a year temporarily reduced the sensitivity of welfare spending to inflation.
- 3.9 The path of RPI inflation largely followed that for CPI, although outturns exceeded our 2010 forecasts by less than for CPI because of the effect of lower-than-expected interest rates on the mortgage interest component of the RPI. This meant that the difference between RPI and CPI inflation – which matters for those costings that involve a switch from one to the other – was smaller than expected. Cumulatively, RPI increased by more than CPI over the forecast period – 19.6 per cent compared to 15.2 per cent – but it only exceeded our 2010 forecasts by around 1 percentage point. The wedge between RPI and CPI inflation over the period was therefore 4.4 percentage points, 2.2 percentage points less than we had forecast at June 2010.

² The accuracy of the full five years of our June 2010 forecast was assessed in our 2016 *Forecast evaluation report*.

Chart 3.1 CPI and RPI inflation from 2010-11 to 2015-16



Note: Outturn here is the rate of inflation in the 12 months to the preceding September, which is the actual rate used for uprating most benefit and tax credit awards, i.e. the 2015-16 data point in the chart is showing the rates of inflation in the 12 months to September 2014. The June 2010 and November 2010 OBR forecasts are based on the rate of inflation from the preceding third quarter, i.e. the 2015-16 data points are showing the rates of inflation in the 12 months to the third quarter of 2014.

Source: ONS

Case study 1: Measures subject to less generous uprating

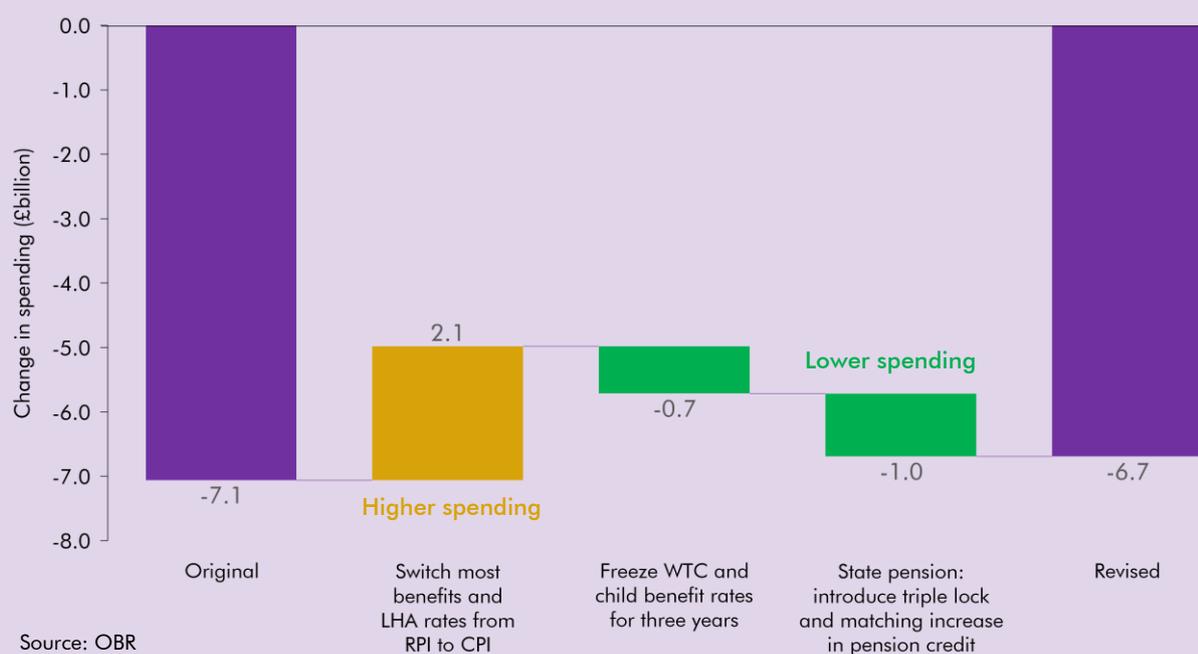
The six measures that comprise this case study are described in Chapter 2. Four of them generate savings by making the respective uprating assumptions less generous – either switching from RPI to CPI inflation (which is typically lower) or simply freezing rates. The other two are the introduction of the state pension triple guarantee ('triple lock') and a pension credit counterpart that matched the state pension cash increase in 2011-12. Taking the six measures in pairs:

- **The two measures that switched from RPI to CPI inflation** (the switching of most benefits to CPI and the introduction of CPI indexation for the local housing allowance (LHA)) generated lower-than-expected savings (a combined £4.2 billion rather than the original £6.2 billion estimate) because the gap between the two different measures of inflation was narrower than we forecast.
- **The two measures that froze rates** (for working tax credit (WTC) and child benefit) raised more than originally expected (£2.5 billion rather than £1.8 billion) because the rate of CPI inflation, which was the baseline for the costings, was higher than forecast in 2010.
- **The introduction of the state pension triple lock** and the accompanying above-indexation increase in the pension credit minimum income guarantee were expected to cost a combined £1.0 billion in 2014-15. Our revised estimate for the triple lock is that it actually reduced spending by £0.6 billion in 2014-15 – while it increased state pension awards relative to both CPI inflation and average earnings, it did not match the growth in RPI inflation (which

was the baseline indexation assumption). The overall cost of the triple-lock exceeded that of its predecessor by 2016-17, and by 2021-22 it had cumulatively cost 4.2 per cent more. By contrast, the pension credit measure cost £0.1 billion more in 2014-15 than expected due to the slightly higher-than-expected RPI-linked increase in state pension awards in 2011-12.

Overall, our revised estimate for these costings is that they reduced spending by £6.7 billion in 2014-15, which is 5 per cent less than the original £7.1 billion estimate (Chart A).

Chart A: Measures subject to less generous uprating: 2014-15 costings



Labour market

3.10 Welfare spending is sensitive to developments in several labour market determinants. Unemployment (and inactivity) affects spending on the out-of-work elements of universal credit (UC) and legacy benefits such as jobseeker's allowance and housing benefit, while employment and average earnings affect spending on the in-work components of UC and legacy tax credits. Chart 3.2 compares our June and November 2010 forecasts of these with the latest ONS outturn estimates for average earnings, employment and unemployment.

3.11 Average earnings growth was consistently weaker than expected in our June 2010 forecast, with a cumulative shortfall between 2009-10 and 2015-16 of almost 10 percentage points. This was related to the similar shortfall in productivity growth relative to forecast.³ Typically, we would expect lower earnings growth to increase spending on earnings-related benefits like working tax credit or in-work housing benefit by keeping more individuals and households within eligibility thresholds. But if earnings growth of benefit claimants were to be stronger than the whole-economy average, then this would not necessarily be the case.⁴

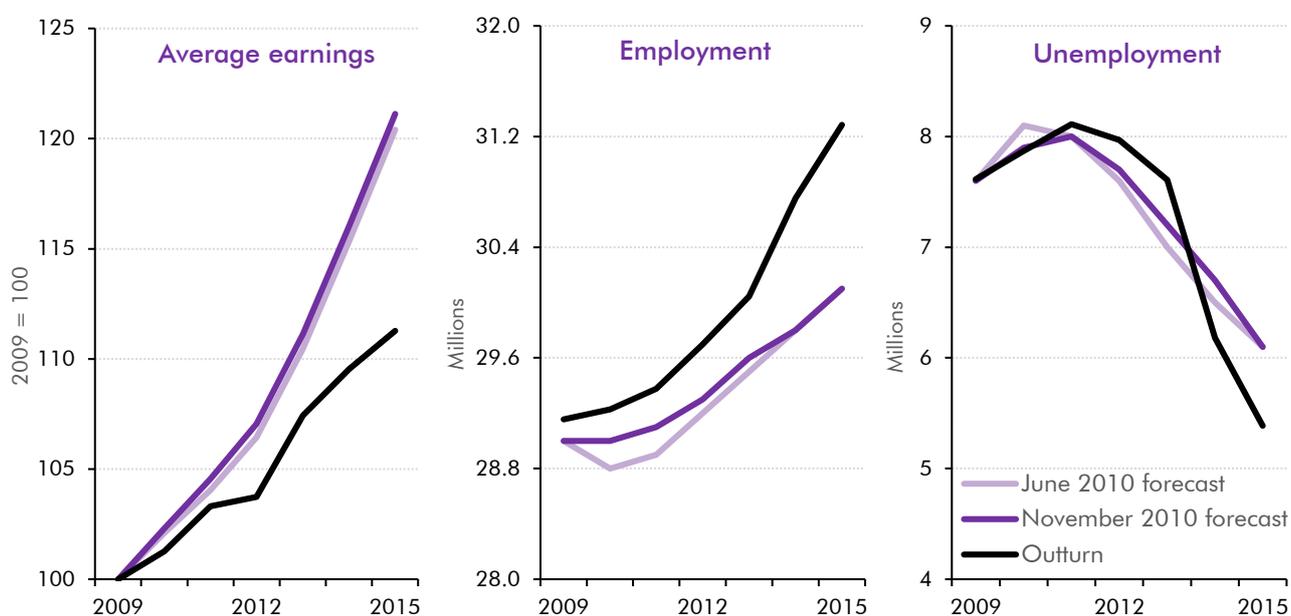
³ For more on this, see our October 2017 *Forecast evaluation report*.

⁴ See Box 4.3 of our March 2018 *EFO* for a discussion of how tax credits claimants' earnings growth at the time appears to have exceeded average earnings growth, putting downward pressure on tax credits spending.

Differences due to baseline forecast

- 3.12 Employment was higher than expected in every year of the forecast. This was driven by a combination of stronger population growth, higher-than-expected labour force participation rates and, by the end of the period, lower-than-expected unemployment. This is likely to have been related to the weakness in average earnings growth, which may have encouraged companies to choose expanding their workforce over capital spending or prompted some people to work more to maintain their income.⁵
- 3.13 Unemployment was initially higher than expected, but then dropped well below our forecast by 2015-16. The sensitivity of welfare spending to unemployment and the wider economy is more limited than is commonly believed and is much smaller than was historically the case, with the largest group of working-age claimants – those on incapacity benefits – showing little sensitivity to the state of the economy.⁶

Chart 3.2 Labour market indicators between 2009 to 2015



Source: ONS

Case study 2: Housing benefit: increasing the age limit for the shared accommodation rate from 25 to 35

This Spending Review 2010 measure, described in Chapter 2, reduced the housing benefit entitlement of single, childless private renters aged between 25 and 35 and claiming under local housing allowance (LHA) rules, from the single-bedroom rate to the lower shared accommodation rate (SAR). It came into effect from April 2012.

The costing used an April 2009 snapshot from DWP administrative data on housing benefit claimants. The static costing simply applied the expected loss in housing benefit per person (an average of around £40 a week) to the baseline number of affected claimants, grown in line with the pre-measures housing benefit forecast (rising to around 100,000 claimants in 2014-15 after

⁵ See our October 2017 *Forecast evaluation report*.

⁶ For more on this, see our May 2022 *Welfare trends report*.

removing those that would have been subject to the SAR regardless). Absent any behavioural response to the policy change, it would therefore have saved £225 million a year by 2014-15.

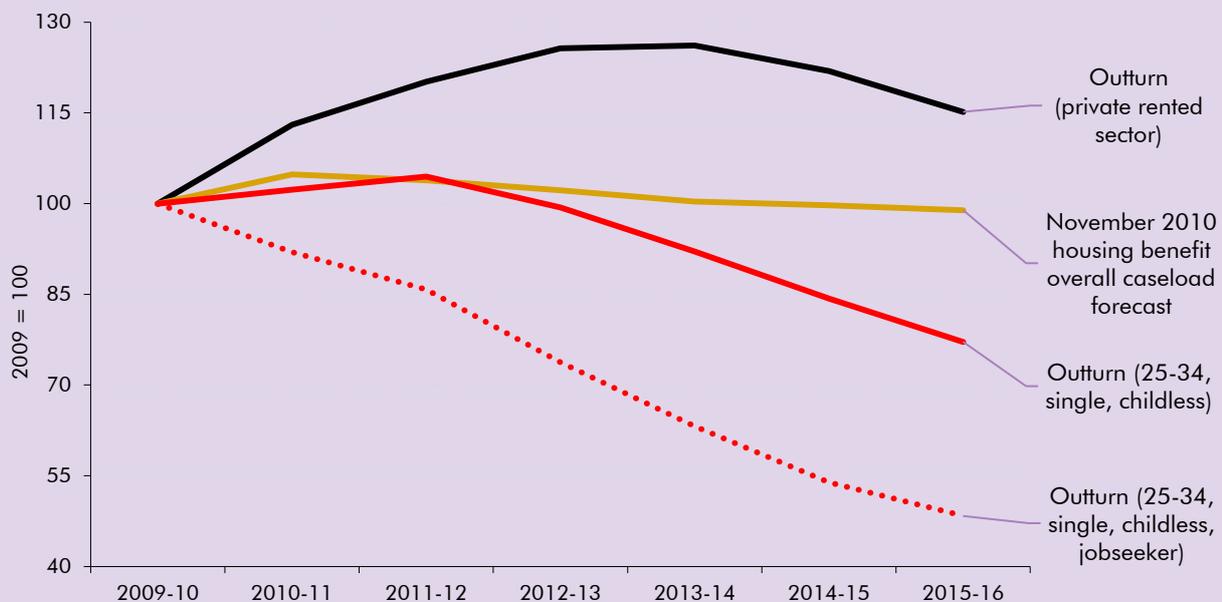
The costing included a small behavioural effect to account for several possible responses: claimants might ask friends, family or partners to move into their household; they may try to qualify for one of the groups exempt from the SAR; some may opt for cheaper accommodation; while others might be dissuaded from living alone to begin with. The costing assumed that the net impact of these was to lower the savings generated by the measure by £10 million a year.

Our revised estimate is that this measure reduced spending by £139 million in 2014-15, 35 per cent less than the original £215 million estimate. The difference can largely be explained by two factors in the costing's baseline that relate to our economy forecast: the stronger-than-expected labour market recovery reducing the caseload, and lower-than-expected growth in rents.

The stronger-than-expected labour market recovery meant a lower-than-expected caseload

Chart B shows that while the growth in the housing benefit caseload in the private rented sector (black line) exceeded our November 2010 forecast for the overall housing benefit caseload (which underpinned the caseload growth used in the original costing, yellow line), growth in the caseload affected by this measure (single, childless claimants in the private-rented sector aged 25-to-34, shown in the red line in Chart B) was much slower. One of the main factors behind this was the stronger-than-expected recovery in the labour market, with employment higher than forecast. This is also borne out by the dashed line in Chart B, which shows an even slower growth rate for the jobseekers' subset of the target group (jobseekers account for a significant minority of housing benefit claimants in the 25-to-34 age range, close to 30 per cent in the April 2009 data).

Chart B: Housing benefit caseload growth



Source: DWP

As a result, the overall SAR caseload was lower than assumed in the costing baseline – around 87,000 in September 2013-14, 71,000 in 2014-15 and 56,000 in 2015-16, so around a quarter

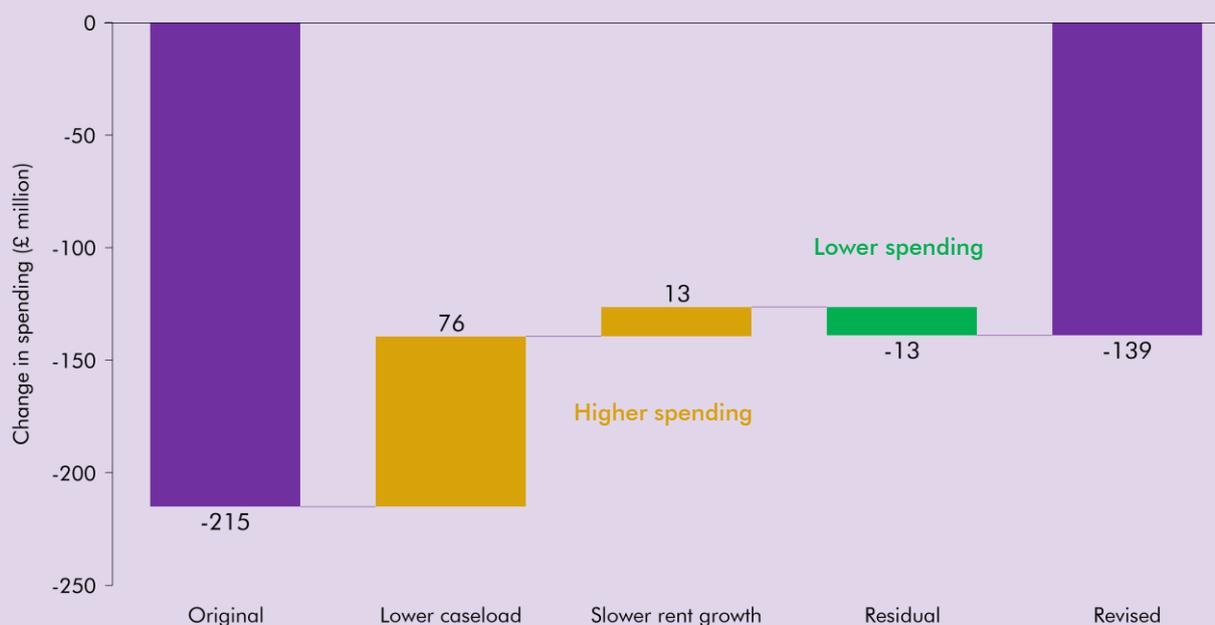
lower. Alongside the performance of the labour market, it is possible that some of this decline is due to a larger-than-expected behavioural response to the measure (or the overall package of measures introduced at the time). One indication of this is that the caseload path among 25-to-34 year-olds went from more closely resembling the 35-to-39 year-old group (pre-measure) to the under 25 year-old group (post-measure). Once we account for the 15 per cent of the 71,000 SAR caseload in 2014-15 who would have received the SAR absent this policy change, our revised estimate of the caseload affected by the policy in 2014-15 is 59,000. Chart C shows that lower caseloads reduce the original costing by £76 million.

The assumed growth rate of local housing allowance rents in the costing baseline was too high

A second, although less important, explanation relates to the baseline forecast for LHA rents, which proved too low. LHA rents at the time of the costing were to be grown in line with CPI inflation from 2013-14, as set out in the LHA package of reforms announced in the June 2010 Budget. This was erroneously omitted from the costing, in which rents were implicitly grown in line with the overall housing benefit forecast, with the most likely explanation for the omission being that it was simply missed given the pace and volume of costings that DWP was producing at the time (the 2010 Spending Review was not accompanied by an OBR forecast and the scrutiny of departmental costings that this involves). In the event, subsequent policies imposed a freeze on rents in 2012-13 and limited growth to 1 per cent in 2014-15 and 2015-16 (these were captured in scorecard costings at the time of announcement). Once we account for the impact of these subsequent policies, we estimate that lower-than-forecast LHA rents account for £13 million of the overall difference in the costing.

The two sources of lower-than-expected savings leave an unexplained residual difference of £13 million. The drivers of the revisions to this costing are summarised in Chart C.

Chart C: Housing benefit shared accommodation rate: revised 2014-15 costing

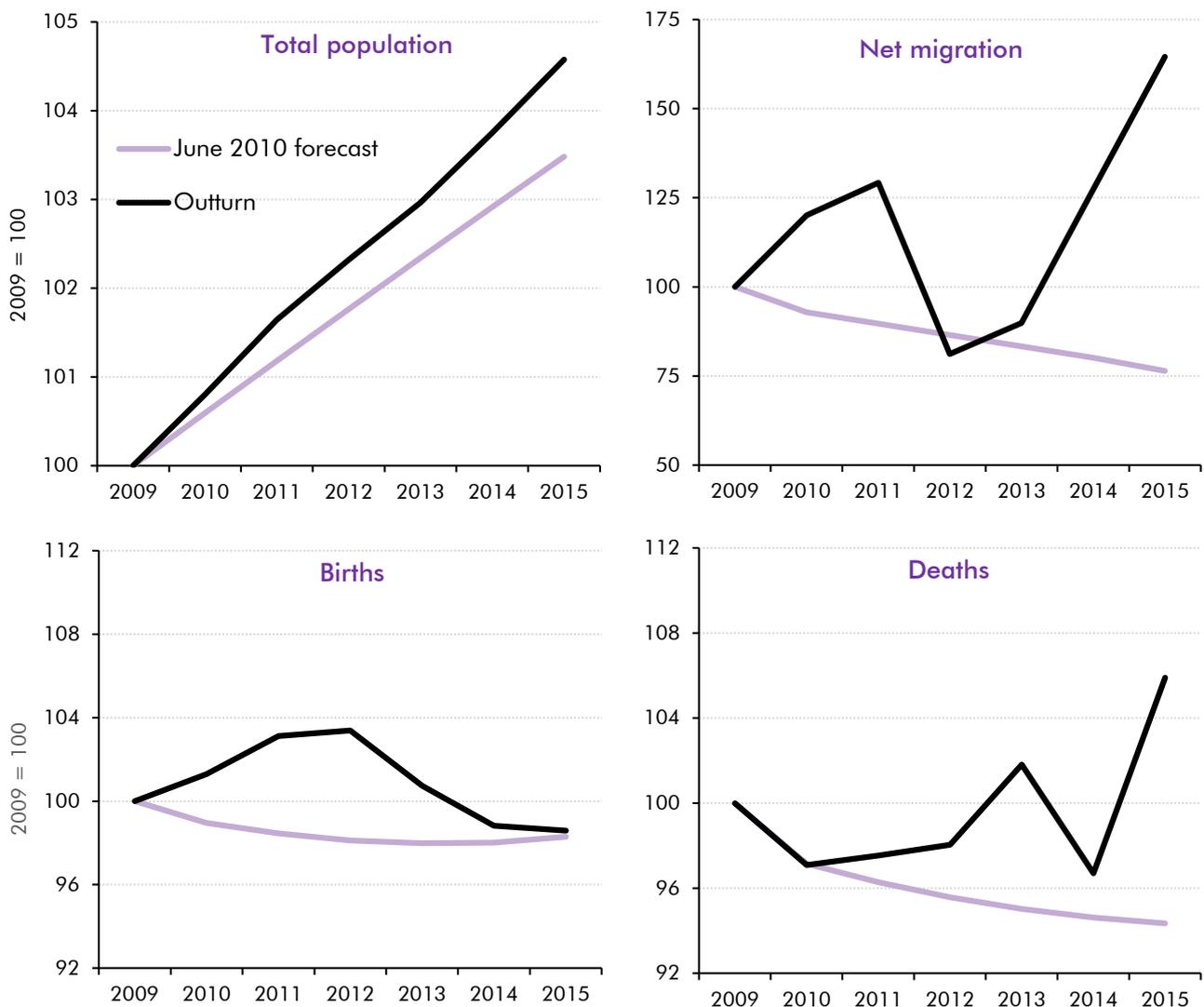


Source: DWP, OBR

Demographic factors affecting welfare spending

3.14 Demographic factors are also an important determinant of welfare spending, which is sensitive to changes in the size and age-composition of the population. For example, the number of births affects child benefit, child tax credit and some universal credit caseloads, while mortality rates and the ageing of the population affect caseloads for the state pension and for other pensioner benefits.

Chart 3.3 Population growth relative to 2008-based ONS population projections



Source: ONS

3.15 Our forecasts in this area rely on the ONS’s population projections, and our 2010 forecasts used their ‘low migration’ variant which predicted net migration of 120,000 by 2015 (the outturn data suggests it was almost triple that amount).⁷ Chart 3.3 shows that population growth was higher than we expected in 2010, growing by a cumulative 4.6 per cent in the five

⁷ The decision to choose the ‘low migration’ variant (see our June 2010 *Pre-Budget forecast*) was in line with a slowdown in net migration in the ONS outturn data from 2007 onwards and reflected our view that the gloomier economic prospects at the time would reduce net inward migration.

years to 2015 compared to the 3.5 per cent growth that our forecast assumed. This was almost entirely explained by higher-than-expected net inward migration, which added around 75 per cent more people to the population between 2009 and 2015 than projected in the 'low migration' variant. The total number of births across the forecast period was 2.4 per cent higher than expected, but this was offset by deaths also being higher, by 3.0 per cent (with deaths in 2015 notably higher than expected due to a more virulent flu strain that year).

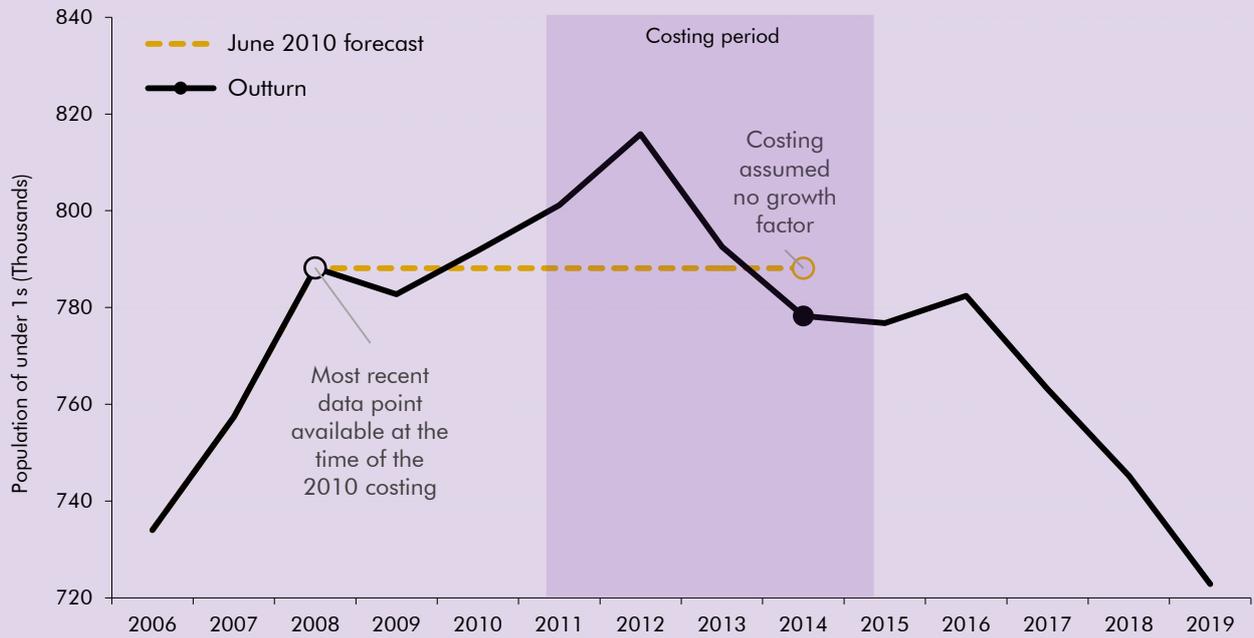
Case study 3: Child tax credit: abolishing the baby element

This measure, described in Chapter 2, removed the additional payment under the child tax credit system for those with a child aged under 1, with effect from April 2011. Both the policy and its costing were relatively simple, with the latter estimated using HMRC's tax credits expenditure and forecast microsimulation model, using lagged, finalised tax credits award data (as interim tax credits data did not contain information on birth rates). The costing was estimated by simply switching off spending related to the £545 annual payment. We assumed no behavioural effects.

The revised estimate for 2014-15 is £230 million, 16 per cent lower than the £275 million original costing, with the entire difference due to a lower-than-expected caseload, which we can explain by splitting into two elements:

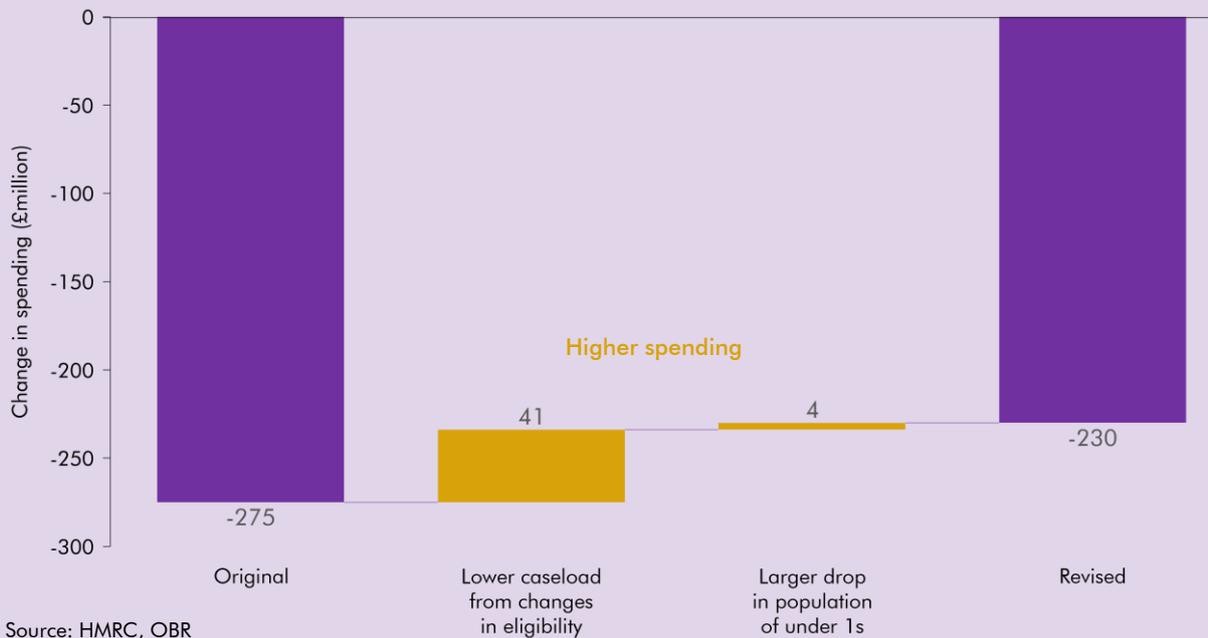
- The most significant is that the **tax credits population turned out to be much smaller than we expected** at the time – by 2014-15 the overall caseload was around 30 per cent smaller than it was in 2008-09, the most recent year of data available at the time of the costing. Changes in eligibility were one of the main reasons behind the decline – for example, mean in-work earnings for tax credits recipients increased by a cumulative 4.3 per cent between 2008-09 and 2014-15, exceeding the 1.2 per cent growth assumed in the costing. This results in lower-than-expected awards for some recipients while removing eligibility entirely from others. Tax credits eligibility was also affected by several other policies that were introduced at the same time as this one, including changes to the income threshold, the taper rate and indexation. The cumulative effect of these lowered eligibility by more than we originally expected and, consequently, reduce the savings from this measure. We estimate that the lower-than-expected tax credits caseload accounts for £41 million (91 per cent) of the overall difference between the original and revised costings (Chart E).
- **A lower-than-expected population of under 1s** (Chart D), which further reduces the savings from this policy relative to the initial costing. Our June 2010 forecast for the baby element did not explicitly apply a growth factor to account for changes in the baby population (though there was a child population adjustment for the child elements of tax credits more generally). This meant that the baby population was broadly held flat for this costing. The outturn data show that the population of under 1s initially rose before falling back from 2011-12 onwards, with a cumulative 1.3 per cent decline between 2009 and 2014, which accounts for £4 million (9 per cent) of the difference in the overall costing (Chart E).

Chart D: Population of under 1s



Source: ONS, OBR

Chart E: Child tax credit baby element: revised 2014-15 costing



Source: HMRC, OBR

Other baseline factors affecting welfare spending

3.16 Alongside the economic and demographic trends discussed above, several other important judgements are required when forecasting the baseline level of welfare spending. Those related to factors within the benefits system include:

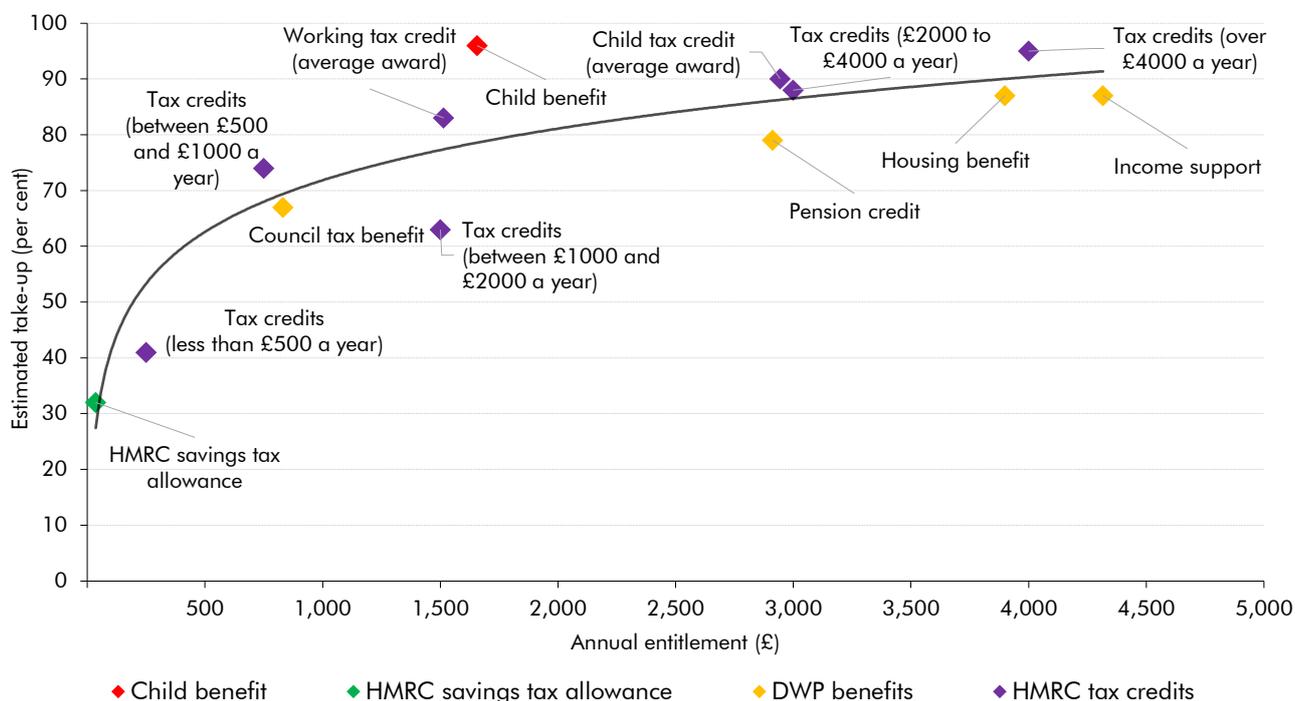
- **Eligibility:** we often need to make judgements about the rates at which individuals and households will be deemed eligible for various benefits. These rates evolve over time with political, operational and legal developments, as well as changes in the composition of the underlying population. We typically compare outturn data with survey data about the characteristics of the underlying population (in particular, DWP's *Family Resources Survey*) in order to inform judgements about trends in eligibility.
- **Take-up rates:** no benefit is taken up by 100 per cent of those who are eligible, and take-up rates vary widely across benefits (Chart 3.4).⁸ When forecasting welfare spending, we typically assume take-up rates remain in line with the latest outturn data, unless there is a clear trend in outturn or policy changes that we expect to change the rate of take-up. It is often difficult to discern whether a change in the caseload for a benefit reflects changes in the eligible population or changes in take-up, so we make judgements about the incidence of benefit receipt in the population, without making explicit judgements about whether eligibility or take-up has been the driver.
- **Delivery:** many of our forecasts require explicit assumptions about the operational delivery of policies or reforms. For example, our forecast for spending on PIP requires assumptions about the volume and speed of reassessments undertaken by DWP's contractors, while our forecast of spending on UC depends on our judgement on the pace at which it will be rolled out across the country and the managed process of moving claimants across from existing benefits. These judgements often require scrutiny of departments' management information, for example on staff numbers and deployment, or the status and capability of IT systems. Many measures announced since 2010 have been subject to operational delays.⁹
- **Fraud and error (F&E):** the latest estimate of the cost of F&E in relation to welfare spending is £6.6 billion in 2021-22 (around 3.0 per cent of spending).¹⁰ As we explained in our March 2022 *Economic and fiscal outlook*, there was a very sharp increase during the pandemic, predominantly related to new UC claims. In the 10 years prior to the pandemic, rates of F&E averaged around half this level. We typically assume, absent policy measures that actively target F&E, that the underlying rates remain constant across the forecast period. There have been several F&E compliance measures in recent years that are now factored into our baseline forecasts, although their impact is small relative to the overall cost of F&E in the system.

⁸ In some cases, measuring ex-post take-up rates can also be challenging.

⁹ See, for example, our 2016 and January 2019 *Welfare trends reports*. Difficulties with operational delivery are not restricted to welfare measures (see Johal, S., *OBR Working Paper No.11: Evaluation of HMRC anti-avoidance and operational measures*, September 2017).

¹⁰ See: DWP, *Fraud and error in the benefit system: financial year ending 2023*; and HMRC, *Child and working tax credits error and fraud statistics 2021 to 2022*. The total cost of F&E represents the 'net' cost of F&E – the difference between F&E that results in lower spending and F&E that results in higher spending – and covers both DWP benefits and HMRC tax credits.

Chart 3.4 Take-up rates for different benefits and tax credits in 2014-15



Note: The 'Tax credits' diamonds refer to a household's overall annual tax credits award.
Source: DWP, HMRC

3.17 We also need to make forecast judgements on some factors associated with, but outside, the benefits system. These include:

- **Childcare costs**, which are a key factor in determining the cost of the childcare element of tax credits and UC. While average current costs can be inferred from administrative data, forecasting requires assumptions about future growth, for which there is little in the way of external estimates for us to draw on.
- **Numbers in full-time education**, which affect eligibility for child benefit, tax credits and UC. Once again, while the Department for Education can provide data on recent outturns, the forecast requires assumptions around future participation rates.

Case study 4: Working tax credit: reducing support towards childcare costs from 80 per cent to 70 per cent

This measure, described in Chapter 2, reduced the amount that eligible households received towards their childcare costs, with effect from April 2011. The costing used HMRC's tax credits expenditure forecast model and lagged administrative data on the childcare element of working tax credit to forecast claimants' total eligible expenditure on childcare. The most recent data available at the time of the costing related to 2008-09.

The static costing simply applied the 10 percentage point reduction to expenditure on the childcare element. It relied on the pre-measure forecasts for the number of families that claim and

the average weekly award. The costing included a behavioural adjustment to allow for some parents switching to informal childcare (which is not eligible for the childcare element).^a

Our current estimate is that the measure reduced spending by £240 million in 2014-15, just under half the original £500 million costing, with the difference largely explained by two differences from the baseline forecast that underpinned the costing, relating to spending on the childcare element and the number of families that would be eligible.

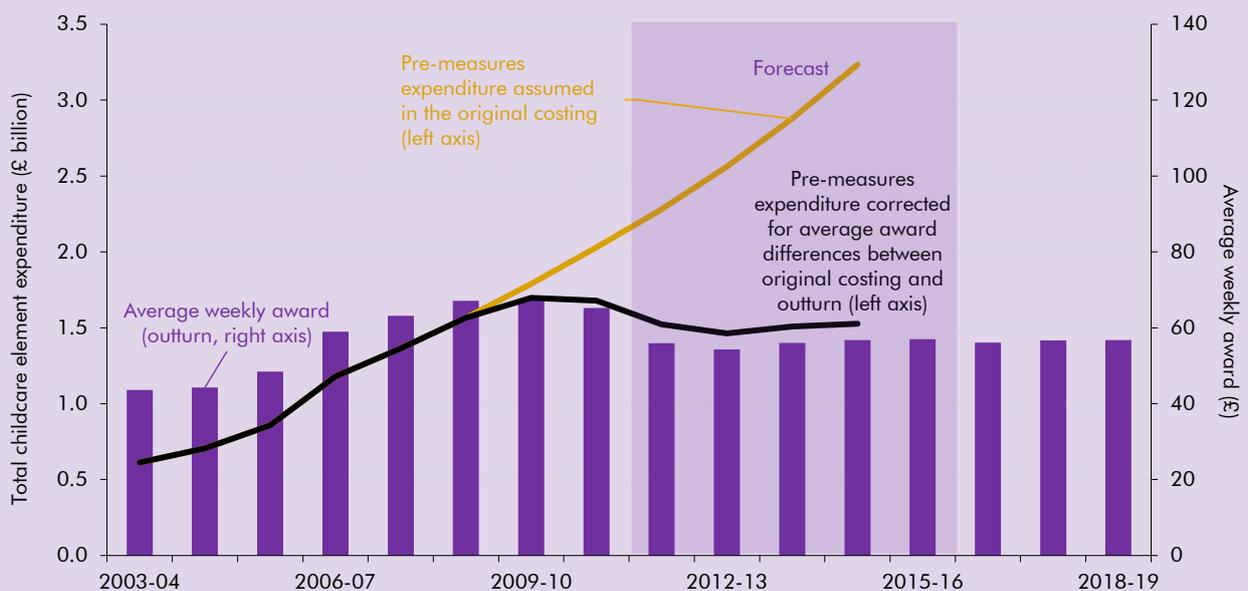
The assumed level of spending on the childcare element in the costing baseline was too high

The costing assumed that, absent the policy, average awards would rise by 7 per cent a year, which was in line with the pre-2008-09 trend but proved to be a significant over-estimate. Chart F shows that underlying childcare costs peaked in 2009-10 and mostly fell thereafter. It also shows a like-for-like comparison between the annual expenditure that was assumed in the original costing (yellow line) and a re-estimated baseline (black line, based on the outturn data abstracting from the effect of the policy) which shows the scale of the difference.

This measure was announced at the 2010 Spending Review, which was not accompanied by an OBR forecast. We subsequently lowered our forecast for average awards at the 2010 Autumn Statement, a month after the policy was announced, from the 7 per cent annual growth assumed in the costing to growing awards in line with earnings (a key driver of childcare costs) which roughly halved the growth rate.

A second factor that led to lower-than-expected expenditure relates to fraud and error, which for the childcare element largely involved claimants receiving higher awards than they should. The costing assumed that the rate of F&E would remain flat over the forecast, but it actually fell sharply, from 25 per cent in 2008-09 to 13 per cent in 2014-15, therefore lowering expenditure relative to expectations at the time of the costing.

Chart F: Pre-measures cost of the childcare element: original vs. latest estimate



Source: HMRC, OBR

The assumed number of families in the costing baseline was also too high

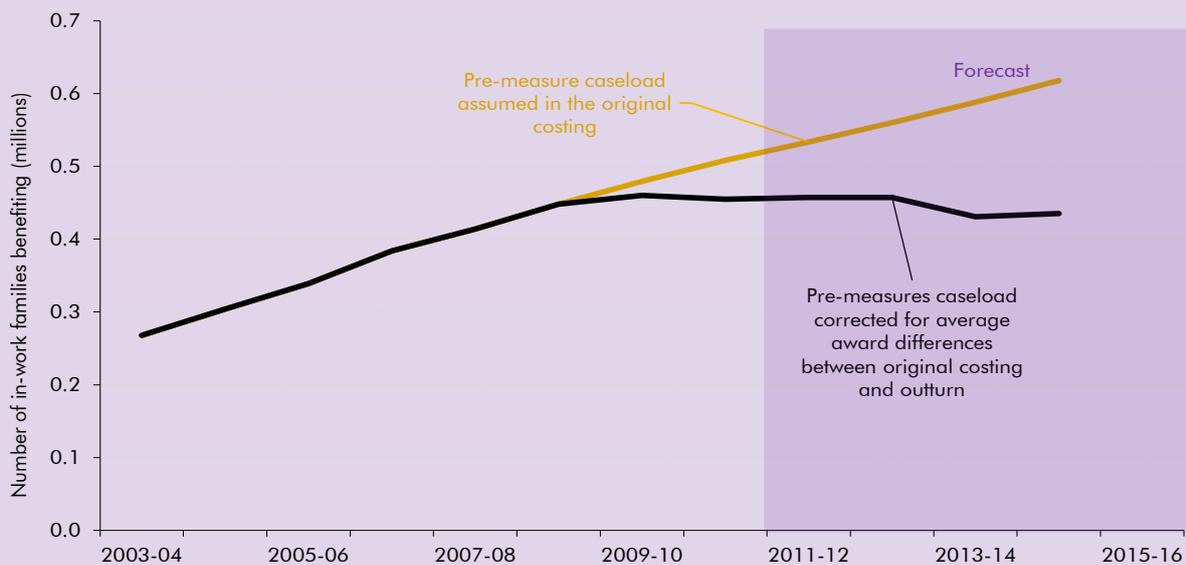
The costing assumed that the pre-measure caseload would increase by 7 per cent in 2009-10, 6 per cent in 2010-12 and 2011-12, and 5 per cent a year thereafter. Chart G compares the baseline caseload assumed in the costing (yellow line) with a re-estimated baseline (black line, once again based on the outturn data abstracting from the impact of the policy). It shows the caseload peaked in 2009-10, perhaps due to the number of families taking up the childcare element reaching a maturation point, before declining thereafter.

Another factor that may have contributed to the caseload decline is the unanticipated fall in the level of female employment during the 2009 to 2012 period, though this had partly reversed by 2014-15. It is also possible that the original costing underestimated the scale of the behavioural response. Again, it is worth noting that at the 2010 Autumn Statement we significantly lowered our forecast of caseload growth to an average of 2.5 per cent a year.

Chart H brings all of these factors together, showing the contribution of each in arriving at our revised £240 million estimate for 2014-15, including:

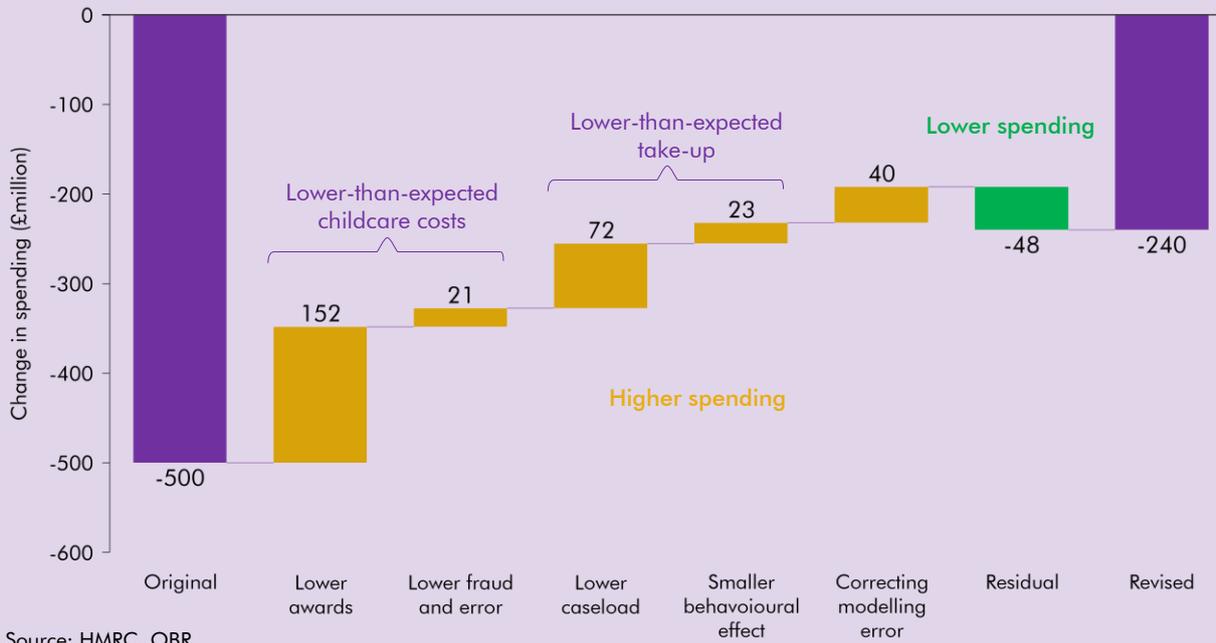
- **Lower-than-expected childcare costs**, which contributed £173 million of the overall £260 million shortfall relative to the costing, of which £21 million is due to lower F&E.
- **Fewer-than-expected families claiming the childcare element**, which contributed a further £96 million. With no evidence to the contrary, we have maintained the proportionate impact of the behavioural response, but the lower static saving means that, in absolute terms, the behavioural impact is now £23 million smaller.
- The analysis for this paper also uncovered **a modelling error in the original costing** that, when removed, reduces the expected saving by a further £40 million.
- These factors combined account for more than the £260 million overall costing difference, meaning there is **an unexplained (positive) residual** of £48 million.

Chart G: Pre-measures number of families: original vs. latest estimate



Source: HMRC, OBR

Chart H: Working tax credit: reducing support towards childcare costs from 80 per cent to 70 per cent: revised 2014-15 costing



^a This saving was informed by estimates in Duncan, A., et al., *Mothers' employment and the use of childcare in the United Kingdom*, October 2001, and the latest available administrative data from 2008-09.

4 Differences within the policy costing

Introduction

- 4.1 In the previous Chapter, we discussed how a costing can deviate from its original estimate by differences arising from the underlying economic and fiscal forecasts. Scrutinising those forecasts is carried out via a separate process, so they are usually taken as an exogenous input that generates the 'pre-measures' or baseline level of spending. In this Chapter, we focus on how costings can under- or over-perform their original estimates due to factors within the policy costing itself.
- 4.2 Most welfare costing begin with an estimate of the '**cost base**', which is the pre-measures level of spending on the population affected by the measure.¹ For a measure affecting a broad population, such as a change in an uprating assumption, the cost base will fall out of the pre-measures forecast, in which case its scrutiny might be limited to ensuring consistency. But when a policy is targeting a narrow subset of claimants the cost base might need to be analytically generated, which might require the use of additional administrative or survey data. In this instance, the estimate of the cost base requires more scrutiny.
- 4.3 Typically, the next stage of a costing involves estimating the '**static**' or **pre-behavioural cost** of the policy change. For welfare measures, that is usually derived by assessing the number of affected claimants and the change in their entitlement. Scrutinising the static costing often involves working through the costing's arithmetic. When the costing is generated from a 'black box' microsimulation model, as is often the case, we use sense-checks to test the plausibility of the estimate, for example multiplying the change in caseload by the change in average award is often a helpful proxy.
- 4.4 The third step in a costing is estimating the **direct or 'first round' behavioural effect** of the measure. This is often one of the more uncertain parts of a costing as its estimate might be based on less reliable data or make use of assumptions. The level of uncertainty is amplified when there are several different behavioural responses within a costing or when there are interactions between several concurrently announced measures. The evidence accumulated from previous measures can be a useful indicator when estimating behavioural responses.
- 4.5 The fourth step is to determine if a measure has an **indirect or 'second round' behavioural effect**. We consider the combined effect on aggregate demand of every measure announced in a fiscal statement using fiscal multipliers. For example, the combined effect of the welfare spending changes announced in a fiscal event affect aggregate demand through our AME multiplier. Welfare policies can also impact the supply-side of the

¹ The 'tax base' is the equivalent baseline level of receipts for tax measures.

economy. For example, we made an upward adjustment to our forecast for medium-term employment due to the policies announced at the March 2023 Budget. Determining the extent that welfare policy changes impact labour participation is another area of uncertainty.

4.6 Our costings scrutiny will also look at a range of other factors specific to each policy, with some of the more common factors being our assessment of whether the relevant department will be able to **operationally deliver the policy as stated**, and how a measure might affect the **level of compliance**. These are additional sources of uncertainty.

4.7 To be transparent about the degree of uncertainty, we assign each certified costing a subjective rating (ranging from 'low' to 'very high') based on our assessment of the uncertainty arising from each of three sources: the data that is used for the costing; the complexity of the modelling; and the scale and complexity of potential behavioural responses to the policy change. We consider the relative importance of each source of uncertainty when determining the overall uncertainty rating.²

4.8 In this chapter we:

- highlight the main **data uncertainties** that arise within welfare costings;
- focus on some of the main **modelling uncertainties**, including the challenges around forecasting the timeliness of operational delivery and dealing with interactions between different measures;
- provide examples of the main **direct behavioural effects** that apply to welfare costings, such as how a policy change might affect the take-up of a benefit;
- discuss the extent to which changes in welfare policy might affect the wider economy, for example by impacting **labour market participation**; and
- use six of the ten **case studies** described in Chapter 2 to illustrate how these different factors have affected welfare costings.

Data uncertainty

Administrative data

4.9 The administrative data that are used in welfare costings are typically of good quality, with both DWP and HMRC administrative systems providing relatively rich and reliable outturn data. Therefore, data uncertainty is usually one of the smaller explanatory factors when

² It is important to emphasise that, when we describe a costing as particularly uncertain, we see risks lying to both sides of what we nonetheless judge to be a reasonable and central estimate. Regardless of the level of uncertainty, we see a 50 per cent chance that the costing is higher than the central estimate, and a 50 per cent chance that it is lower. Others may differ in their views of the central estimate or the uncertainty around it, with the scope for different views likely to be greater for more uncertain costings. See our online '*Policy measures uncertainty database*' for our measure-by-measure ratings at the time each was announced.

assessing the overall level of uncertainty in welfare costings that use administrative data. There are two main uncertainties relating to administrative data:

- Its **timeliness relative to the start date of the policy change** that is being costed. For example, HMRC's tax credits 'finalised' award administrative data are only available after a two-year lag, meaning that costings using this data must decide between using final data that is somewhat out-of-date or more up-to-date but provisional data (such as HMRC's 'interim' tax credits data, which uses a snapshot of the tax credits system that is produced on a quarterly basis). In both cases the lack of *timely* administrative data introduces an element of uncertainty even when the quality of the data is high.
- The information that is recorded in administrative data is still subject to **reporting and measurement error** – for example, the regular under- and over-payment of awards is an established characteristic of the benefit system.³ One of the issues raised by under- and over-payments (and their causes) is that the underlying data might not be fully representative of the population relevant to the costing. For example, if a payment in arrears is recorded at the point the payment is made rather than when it accrues; or when an administrative backlog results in a payment being recorded in a later period than it should. The most recent published statistics show that total under-payments in 2020-21 amounted to £2.7 billion, while over-payments amounted to £9.4 billion.⁴

Case study 5: Working tax credit: increase the working hours requirement for couples with children from 16 to 24 hours a week

This Spending Review 2010 measure, described in Chapter 2, raised the number of hours that couples with children need to work each week to meet working tax credit eligibility requirements. It came into effect from April 2012.

The costing used a snapshot of tax credits administrative data from August 2010 and projected it forward using HMRC's tax credits expenditure forecast model. This 'interim' data provided more up-to-date information on claimants' working hours than the available finalised data, but with the additional uncertainty associated with it being provisional.

The static costing simulated the reduction in entitlement for existing tax credits households whose recorded hours failed to meet the new, higher, threshold. It also assumed that households' working hours would, between 2012-13 and 2014-15, return to the higher rates observed in the pre-recession period (therefore higher than the hours observed in the post-recession base data). An adjustment was made for assumed under-reporting of hours worked in the administrative data, but the costing did not factor in any explicit assumptions about behavioural responses over and above this.

³ It is important to remember that payment in arrears does not just arise due to F&E. For example, both employment and support allowance and personal independence payment allow individuals to challenge decisions made by DWP regarding their benefit claims (a *mandatory reconsideration* is the first step in that challenge while *appeals* can be made if a claimant is dissatisfied after the mandatory reconsideration).

⁴ Between 2010-11 and 2019-20 under-payments averaged £5.0 billion a year while over-payments averaged £1.8 billion, so the pandemic-affected 2020-21 figures are markedly higher (see our March 2022 EFO).

The measure was originally expected to reduce spending by £535 million a year from 2012-13 onwards.^a The revised costing suggests that savings were only around a quarter of that estimate, largely due to a higher-than-expected post-measure tax credits caseload, with more households meeting the 24-hour threshold than assumed in the costing. Our analysis identifies two main contributory factors: the under-reporting of hours worked in the administrative data, and higher-than-expected post-measure working hours.^b

The pre-measure estimate of the number of hours worked by existing households was too low

The administrative data that fed into the costing was based on households only needing to demonstrate that they met the minimum 16-hour (pre-measure) eligibility threshold – they did not have to specify the precise number of hours that they worked. There was, therefore, bunching in the administrative data at the 16-hour threshold – some proportion of which would be due to ‘under-reporting’ of hours relative to the true position. This was a key source of uncertainty in the costing, which included a simple adjustment to try to account for it. More precisely, the adjustment aimed to capture the fact that some households meeting the previous 16-hour requirement would (if true hours worked had been recorded) also meet the new 24-hour requirement.

Drawing on operational intelligence within HMRC, it was assumed that 2.5 per cent of affected households were under-reporting hours in this way. The outturn data suggest this uplift was insufficient, with just over half of the 200,000 households identified as working between 16 and 23 hours in April 2012 reporting increases taking them above the 24-hour threshold within the first year of the new rules.

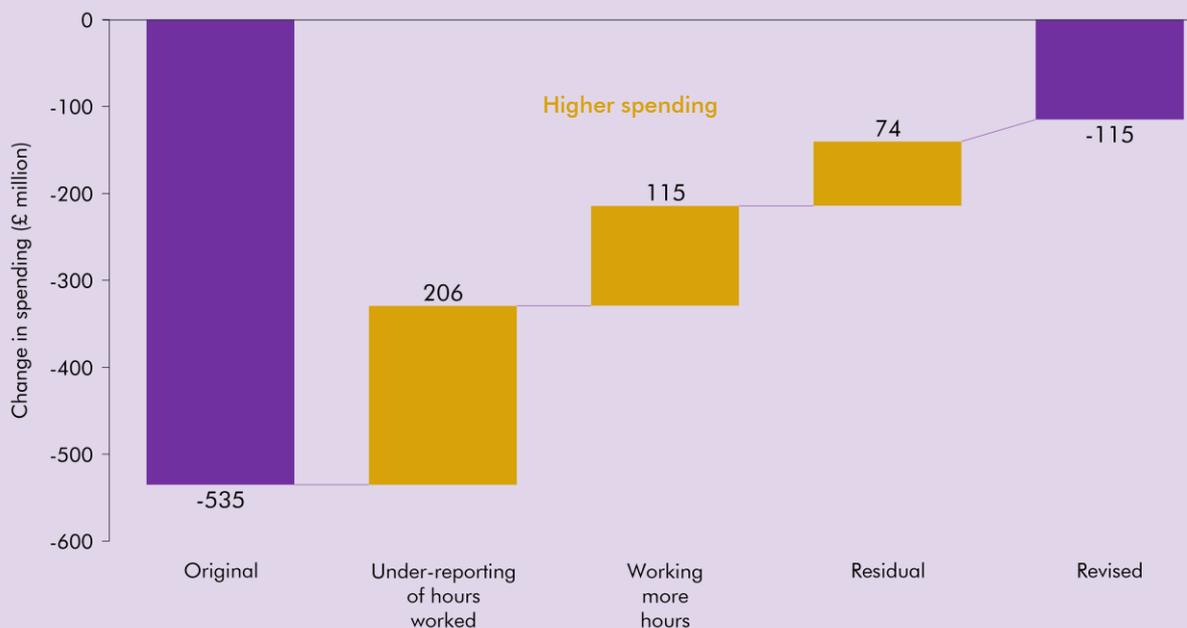
We estimate that the under-reporting of hours worked in the pre-measure administrative data is responsible for £206 million (49 per cent) of the overall difference between the original and revised 2014-15 costing (Chart A).

Households’ working hours rose by more than we expected

While we think that the under-reporting of hours is the largest single factor in explaining the higher-than-expected post-measure caseload, we estimate that there was also a sizeable contribution from those households that chose to increase their hours worked to meet the new rules – a behavioural response that was not explicitly included in the original costing. Analysis of the interim data over the period suggests that 75 per cent of those reporting higher hours also reported higher earnings, implying a genuine increase in hours worked (supported by the wider employment recovery described in Chapter 3).

We estimate that households responding to the new rules by working more hours explains £115 million (27 per cent) of the overall difference between the original and revised 2014-15 costing (Chart A). This reflects the welfare spending cost of retaining eligibility for working tax credit by increasing hours worked. It does not attempt to capture the impact on tax receipts from this income growth, which would be modest for the households involved due to the income tax personal allowance and more modest still at the whole economy level if the additional hours worked by these households displaced those worked by others not affected by the measure.

Chart A: Working tax credit: increase the working hours requirement for couples with children from 16 to 24 hours: revised 2014-15 costing



Source: HMRC, OBR

This leaves an unexplained residual of £74 million, which is likely to reflect several factors, including some uncertainty in the outturn data, the challenge of isolating the impacts of this measure from the large number of others announced during the same period and the difficulty in distinguishing the effects of measures from the underlying surprises observed in the wider economy and their impacts on our pre-measures forecast (described in Chapter 3).

^a Under the accounting treatment at the time this was split into £430 million a year from lower spending on tax credits and £105 million a year from lower negative taxes.

^b We do not think that higher fraud and error in the reporting of hours was a factor.

Survey data

4.10 Where administrative data are not available or appropriate, **survey data** can serve as a useful substitute. DWP's principal model for welfare costings, its policy simulation model (PSM), is based on the survey estimates of characteristics in the underlying population drawn from the annual *Family Resources Survey* (FRS). Detailed surveys like the FRS provide a wider range of information about benefit recipients and, importantly when considering a policy change that would affect people not currently in receipt of a benefit, also cover individuals and households that are not in contact with the administrative systems.

4.11 There are three main drawbacks when using the FRS:

- The **relatively low sample sizes** can introduce uncertainty around how representative the results are when scaled up to the whole population. For the survey as a whole, this is not an issue, with a sample size of around 16,000 for the most recent FRS. But this source of uncertainty can become material when the data must be disaggregated for

particular policy costings that need to drill down to individual characteristics within a single benefit, where the relevant sample size might be very small and, therefore, the resulting conclusions less robust.⁵

- The **survey results are produced with a two (or sometimes three) year lag**, while the time needed to incorporate the new data into costings models adds a further delay. This means that the models must initially project forward from the most recent sampling year to the point at which the costing takes effect. For example, costings at Spring Budget 2023 used FRS data for 2019-20, a minimum of four years earlier than the first year of the costing.⁶
- The FRS is also subject to **measurement error**. Indeed, the shortcomings of self-reported survey data – such as the accuracy of interviewers’ questions and respondents’ answers – are well known.⁷ For example the FRS has a variable that captures an individual’s interpretation of their ‘usual pay’, which might not be a close match to the specific earnings measure that is required for a costing. Surveys are also subject to bias in the distribution of those that choose not to respond, a sub-group that is unlikely to resemble a random sample due to the likelihood that its members share certain common characteristics that lead them not to respond to the survey.⁸

Modelling uncertainty

4.12 Policy costings are generated using many models that are owned and maintained by the department responsible for the measure. Most costings in this paper are based on one of two models:

- **DWP’s policy simulation model (PSM)** is based on the annual FRS, which is discussed above. The time lag associated with the FRS and its eventual incorporation into the PSM means that the modelling of costings may not capture recent developments in the economy nor the impacts of new policy announcements. If these effects must be captured via other means – for example by calibrating to alternative estimates or using off-model adjustments – then while this process can be effective at aligning the older data with the emerging data, it does add a further layer of uncertainty. The PSM is only used for costings, which opens us the possibility that some results might not be consistent with the relevant forecasting model.
- **HMRC’s tax credits expenditure forecast model (TCEFM)**, which uses administrative data from a very large random sample of around 500,000 tax credits claimants. This

⁵ Box 4.1 in our January 2018 *Welfare trends report* (that focuses on universal credit) uses 2015-16 FRS data to highlight this issue, saying that there are “around 25 benefit units for the [working tax credits]-only, non-self-employed, no housing benefit, single’ calibration group, which is expected to account for just over 170,000 cases in the steady-state UC population”.

⁶ This is an unusually long delay due to pandemic-related factors. DWP chose to use an earlier but higher-quality vintage of data (with a larger sample size) than more recent but lower-quality data collected during the pandemic – a choice that we scrutinised and certified.

⁷ For more on the FRS, refer to DWP, *Family Resources Survey: background information and methodology*.

⁸ Costing models, such as DWP’s policy simulation model, can correct some of the impact of the measurement error by grossing-up to administrative data caseloads and forecasts. This process can be effective at dealing with the issues raised by measurement error, especially at an aggregate level. But it might be less effective for those costings targeting a smaller sub-group, unless the grossing is done at the administrative level of the sub-group affected by the measure.

reduces sampling error in the modelling of claimants' tax credits entitlement under different policy scenarios. The TCEFM is also used to forecast overall tax credits expenditure, ensuring consistency between forecasts and costings.

- 4.13 Both the PSM and TCEFM are examples of microsimulation models that use detailed micro-level data to recreate the benefits system and estimate the effect of policy changes at the individual and household levels, before aggregating and scaling up to obtain the overall effect of a policy change. Some of the benefits of microsimulation models include:
- They are usually **better at identifying the impact of policies on different types of individuals** in the population than top-down models that are based on aggregate data.
 - More granular information allows for the **simultaneous modelling of entitlement to multiple benefits** and may capture the impact that a change in one benefit has on others, highlighting potential knock-on impacts or unintended consequences.
 - Being able to assess the parameters that influence individual responses to policy changes allows a **better understanding of potential behavioural effects**, especially where these vary across the population.
 - The **flexibility to extend or amend the model**, for example to account for changing characteristics of claimants.
- 4.14 Microsimulation models also have their drawbacks. In addition to those associated with data uncertainty that are discussed above, some of those specific to the PSM and TCEFM include:
- Both models are static rather than dynamic, in that individual and household **characteristics do not change through time**. This means that the underlying assumptions about the population may become increasingly unrepresentative the further into the future the model is being used, particularly when the policy environment is changing. Representativeness can be improved by calibrating the sample weights associated with household characteristics, but any adjustments will inevitably add a further layer of uncertainty.
 - Both models are empirical rather than theoretical models, meaning that they **do not attempt to explain why things change** but rather highlight possible channels. This is an advantage in the sense that the model accurately reflects what has happened in the past, and of course the models are effective in tracing back the impacts on individual benefit streams from a change in a policy parameter. But it can be problematic if past relationships change, for example due to changes in the economy, and the model is unable to adapt.⁹ Deviations from previously observed patterns may result in an observed change being incorrectly assigned to the wrong underlying factor, which would

⁹ Of course, forecasting models with a stronger theoretical basis can still struggle when faced with economic shocks or other unprecedented changes. It is also important to note that analysts in DWP and HMRC build and adapt their models, partly to improve forecast performance.

adversely affect any policy costing that includes that factor as an important component (or omits a now relevant factor because it was not previously prominent in the data).¹⁰

- More generally, while microsimulation models can provide a richness of analysis unavailable to aggregate models, this **often comes alongside higher running costs**. The underlying data needs to be acquired, processed and maintained and a team of experts will be required to build, maintain and operate the models. The models might also be resource-intensive in terms of the computational power needed to simulate large populations or detailed individual-level interactions. From an OBR perspective, it is also worth emphasising that the outputs from microsimulation models **can be less transparent**, especially since we have limited access to the underlying data, which is governed by rules around benefit-recipient confidentiality.¹¹

Uncertainties around operational delivery

4.15 Modelling uncertainty can arise from the time it takes departments to implement policies that include operational changes, such as introducing new IT systems, recruiting and training additional staff, or outsourcing functions to external contractors. We have previously noted that activities that rely on successful operational delivery have often taken longer and yielded less than originally expected. Indeed, ensuring such costings include a credible implementation plan that is based on evidence from previous operations and allows for sufficient contingency is a key lesson the OBR has acted on since 2010.¹²

4.16 Forecasting the implementation of a major reform or introduction of a new benefit is especially challenging since it adds two further sources of modelling uncertainty:

- For a wholly new benefit (or one that replaces an existing one) there may be **insufficient data to determine a reliable pre-measure baseline**. To get around this, the costing may then need to construct a counterfactual baseline that relies on several uncertain modelling steps – for example, for UC this involved constructing both ‘no-UC’ and ‘full-UC’ counterfactuals, both of which were based on limited information.¹³
- Large-scale changes due to be implemented over several years are also subject to **policy risk** – both in terms of subsequent changes in policy design and decisions to slow the pace of rollouts.¹⁴

4.17 The current end dates for the rollout of UC and PIP – the two major reforms to the working-age welfare system that were announced in 2010, the year the OBR was formed – are both beyond our current forecast horizon, which runs to March 2028. At best, this leaves them 11 and 12 years behind schedule, respectively.

¹⁰ This is true of other models too, but the issue will tend to be more acute with microsimulation models that are based largely on purely empirical relationships and exacerbated by the extended lag between the base year of the data and the first year of the policy change.

¹¹ There are similar rules around taxpayer confidentiality for tax costings.

¹² For example, see our 2017 *Fiscal risks report* and our *Welfare trends reports* from January 2018 (on universal credit) and January 2019 (on personal independence payment) respectively. Slower-than-expected operational delivery is also a characteristic we noted in tax policy costings, see Johal, S. (2017), *Evaluation of HMRC anti-avoidance and operational measures*, September 2017.

¹³ See our January 2018 *Welfare trends report*.

¹⁴ In both cases the fiscal impacts would be captured in our forecast as a new policy costing

Case study 6: Contributory employment and support allowance: introduce a one-year time limit for those in the work-related activity group

This Spending Review 2010 measure, due to begin in April 2012, changed the rules for claimants of contributory employment and support allowance (ESA) that are placed within the 'work-related activity group' (WRAG). Under the new rules, contributory ESA was to be paid for a maximum of one year for those in the WRAG and would cease thereafter. Those in the 'support group' (with more severe health conditions) would continue to receive payments for as long as they met the eligibility criteria.^a

The original costing, estimated using DWP's policy simulation model (PSM) and based on 2007-08 data from the Family Resources Survey, relied on some highly uncertain assumptions.^b Notably, ESA was only introduced in 2008-09 so the FRS data underpinning the costing did not include any actual ESA claimants. Therefore, even establishing the pre-measures forecast to which the policy change would apply was challenging as it reflected survey responses from claimants of ESA's predecessors, such as incapacity benefit (IB) and severe disablement allowance.

The PSM used these proxy cases to calculate the change in ESA entitlement for those affected by the measure, as well as the knock-on impacts on other benefits. The costing also considered several uncertain behavioural responses, including claimants: reducing their income (and/or drawing down savings) to qualify for income-based ESA; claiming other benefits; appealing the outcome of unfavourable work capability assessments (WCAs); and moving into work.^c

Our current estimate is that the measure reduced spending in 2014-15 by £0.4 billion, nearly 80 per cent less than the original £2.0 billion estimate. Several factors contributed to this material shortfall, which is the largest among our case studies. These include: revisions to the pre-measures forecast; a smaller-than-assumed proportion of new ESA claimants flowing into both contributory ESA and the WRAG; a smaller-than-assumed proportion of claimants of ESA's predecessor, IB, being placed in the WRAG (with more placed in the support group); a higher-than-assumed proportion of claimants moving onto other benefits; and over-estimating the pace of the overall ESA rollout.

The pre-measure forecast underpinning the costing was too high

Measures that are announced in fiscal statements are accompanied by a pre-measures OBR forecast that provides the baseline level of spending for costings.^d There was no OBR forecast alongside Spending Review 2010, so this costing used the OBR's June 2010 Budget forecast as a pre-measures baseline instead. That forecast, our first, overstated spending on ESA to such an extent that it was revised down by 23 per cent in 2014-15 in our November 2010 forecast.

The revision made use of the administrative data emerging during 2010, including on WCA success rates. The introduction of ESA had led to a substantial increase in appeals, resulting in a backlog that meant the final outcomes of assessments were slow to emerge. Initial findings were available for the costing, but with relatively few recorded claimants the data were unlikely to be representative of the wider population, reducing their reliability as an alternative or secondary information source.

The uncertainty around the pre-measure level of expenditure on ESA was compounded by inconsistencies between the PSM and the ESA forecasting model, with the former based on overall

ESA spending whereas the latter split out spending between the WRAG and support group. One lasting lesson from this is that, whenever possible, costings should be accompanied by a full OBR forecast. That would have allowed the PSM results to be calibrated to the forecast model.

Chart B shows that revising the pre-measures forecast accounts for £0.2 billion (14 per cent) of the overall £1.5 billion difference between the original and revised costing.

Fewer new cases than assumed in the original costing

One implication of using proxy data was the need for assumptions to be made in the modelling. Chart B shows that our latest view of two such assumptions, which estimated the volume of new cases affected by the measure, together contributed £0.7 billion (45 per cent) of the overall difference between the original and revised costing:

- The proportion of new ESA cases claiming contributory ESA was expected to be 32 per cent in 2014-15 but only reached 25 per cent, with the difference accounting for £0.3 billion (18 per cent) of the overall difference between the original and revised costing.
- The costing assumed that 75 per cent of post-WCA new cases would enter the WRAG, leaving 25 per cent entering the support group, but the actual WRAG share was around 40 per cent by 2014-15. This was partly due to the underperformance of the contractors hired by DWP to deliver WCAs, which led to prioritisation of support group recommendations at the expense of WRAG cases. The associated backlogs also slowed the pace of IB reassessments and the subsequent migration of IB cases to ESA. Furthermore, a series of internal and external reviews made recommendations (accepted by DWP) that increased the proportion of cases moving into the support group. This element contributed £0.4 billion (27 per cent) of the difference between the original and revised costing.

Incapacity benefit reassessments led to a lower-than-expected WRAG caseload

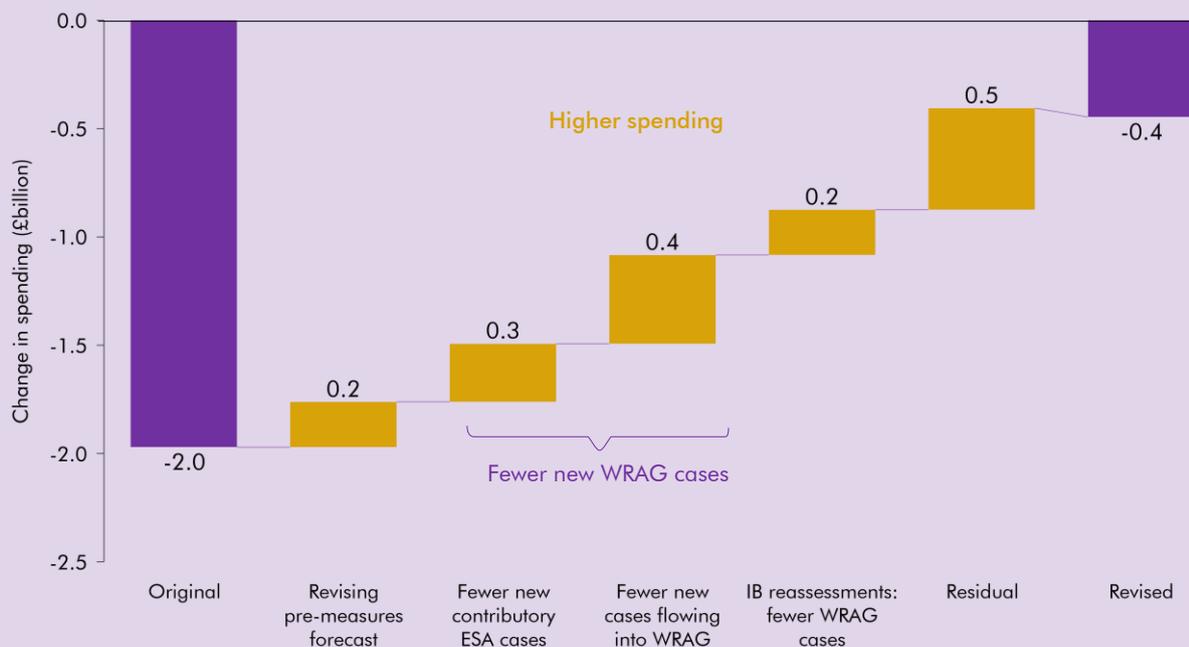
The costing included an allowance for the impact of IB reassessments, which applied a WCA to determine whether IB recipients qualified for ESA, and which group they were placed into if they did. The outturn data suggest that fewer IB cases than expected entered the WRAG, with more entering the (more costly) support group. Issues with contractor performance also meant that far fewer reassessments than expected took place. Chart B shows that this element accounted for £0.2 billion (14 per cent) of the overall shortfall in the costing.

Higher-than-assumed numbers of affected claimants remaining in the benefits system, and the slower rollout of ESA more generally, also reduced savings

Around £0.5 billion (31 per cent) of the difference from the original costing is contained within the residual (Chart B). One likely component of this is that more cases that were moved off ESA due to the measure subsequently remained in or returned to the benefits system than allowed for in the costing. The original costing estimated that just over 50 per cent of the static savings from the measure would be offset by claimants either re-claiming ESA or moving to other benefits. These claimants have not been tracked so we cannot quantify the true figure, but it seems likely that the size of the offset should have been greater. Another element within the residual is the slower-than-assumed pace of the full ESA rollout, with fewer ESA cases reducing the potential

savings from this measure within its original timeframe. The fact that the roll-out of new benefits frequently happens more slowly than initially assumed is another lesson we have taken on since 2010, with larger delays witnessed in the rollout of UC and PIP than was the case with ESA.

Chart B: Contributory employment and support allowance: introduce a one-year time limit for those in the work-related activity group: revised 2014-15 costing



Source: DWP, OBR

Overall, the outturn data show that there were around 750,000 contributory ESA claimants in 2014-15, 40 per cent fewer than assumed, with only 35,000 of those in the WRAG (contributing £0.2 billion of expenditure). It also shows that this was largely offset by more claimants for income-related ESA, around 1.3 million in 2014-15, 49 per cent higher than expected, including just over 400,000 in the WRAG (£2.6 billion of ESA spending). The outturn data also show that the numbers in the costlier support group exceeded expectations.

^a All ESA claimants (contributory and income-related) must complete a 'work capability assessment' (WCA), which determines how their illness or disability affects their ability to work. If the WCA deems that a claimant has some, limited, capability for work then they are placed in the WRAG, where they receive regular work-related support, including interviews with an adviser. If the WCA determines that a claimant has a severe health condition, making it unlikely they will be able to work in the foreseeable future, then they are placed in the 'support group', where there are no work-related requirements. As well as those in the support group, those in the WRAG that were exclusively claiming income-related ESA (the means-tested element) were also excluded from the new time limit. In practice, those claiming both contributory and income-related ESA were also excluded from the time limit. For existing WRAG claimants, the new limit applied once they had reached a year in the group, so for those that were already beyond the one-year limit, the effect was immediate. Some claimants affected by the measure might, subsequently, still qualify for income-related ESA.

^b See Chapter 3 for more on both the policy simulation model and the Family Resources Survey.

^c The first three of these would increase welfare spending while the last would reduce it.

^d Or the baseline level of receipts for tax measures.

Interactions between different costings and benefits

4.18 Measures that are announced simultaneously will often interact with each other and, indeed, the structure of the benefits system links entitlement across the different benefits. These interactions need to be accounted for if we are to capture the full impact on the public

finances, though how they are captured can be a presentational choice for the Treasury. A good example is the benefit cap, which was introduced at Autumn Statement 2010 and whose monetary value was reduced in the July 2015 Budget. The cap applies to aggregate entitlement so interacts across the benefits system. This in turn means that the cost or yield from changing it can be affected by whether the interactions between it and other measures are reflected in benefit cap costings or other costings – something that is determined by the order that measures appear on the Treasury’s policy ‘scorecard’. For example, in the July 2015 scorecard, the benefit cap was placed ahead of several measures designed to reduce welfare spending – this raised the estimated savings from the cap but reduced savings from the subsequent measures, with no net impact overall.¹⁵

- 4.19 Interactions may also come outside of the benefits system. For example, the size of awards often depends on the level of a claimant’s post-tax income, so changes to the income tax system often affect welfare spending, such as the knock-on effects from changes to the personal allowance on UC awards.

Case study 7: Benefit cap

This Spending Review 2010 measure introduced, from April 2013, a cap on the total amount of payments a single household could receive from certain out-of-work benefits.^a The cap was set at £500 a week for couples and lone parents, and £350 a week for singles, and was applied by local authorities, who are responsible for administering housing benefit, reducing capped households’ housing benefit awards. These limits were held constant in cash terms, therefore becoming more restrictive since benefit awards are typically uprated each year (although many were subject to freezes in the mid-2010s).

Households entitled to working tax credit were exempt from the cap, as were those claiming certain disability and incapacity benefits. Some of these exemptions reflected amendments to the original policy before it was introduced, such as those announced at Budget 2012.^b

The original 2010 policy was forecast to save £270 million in 2014-15, but this was revised to £265 million once the pre-implementation amendments were factored in (and it is this latter estimate that we focus on).^c The costing was calculated using DWP’s policy simulation model populated by lagged Family Resources Survey data, with the savings estimated by simulating the reduction in entitlement for affected households.

Importantly, the costing is only capturing the total amount of benefit payments that would be capped, known as the static costing, and did not include any impacts from a range of potential behavioural responses (that we discuss below). The justification for excluding these was the belief that the static costing was a central, albeit uncertain, estimate – it was deemed that there was an absence of reliable evidence to guide any assumptions on the scale of the behavioural responses but also that there was uncertainty in both directions.

¹⁵ Placing the benefit cap after all the other measures on the scorecard would have had the opposite effect – raising the savings from each preceding measure and reducing those from the cap. In our July 2015 *EFO* we showed that the estimated savings from the benefit cap (at the time) would have been 60 per cent lower if it had been placed last on the scorecard.

Our revised estimate is that £80 million of payments were capped in 2014-15, 70 per cent lower than the costing. The main factor behind the difference is a shortfall in the number of households affected, with a smaller impact from lower-than-expected amounts lost per case.

Fewer households were affected by the cap

Administrative data show that an average of around 25,000 households a month were capped during 2014-15. This compares to around 58,000 assumed in the costing. There are several potential causes for this shortfall, including several possible behavioural responses:

- **Moving into work.** If those affected responded by moving into work then this would cut welfare spending on out-of-work benefits, though it would be likely to increase spending for some on in-work welfare payments such as tax credits. The evidence from a 2014 DWP evaluation suggests that capped households were around 5 percentage points more likely to move into work than similar uncapped households, amounting to around 2,500 cases.^d
- **Moving to a home with a lower rent.** In the same evaluation report, DWP noted a modest increase in the number of households moving within a year, largely among those who stood to lose larger amounts as a result of the benefit cap.
- **Higher-than-expected take-up of the exempt benefits.** A DWP survey noted that around 10 per cent of initially capped households moved to an exempt benefit. It seems likely that, at least in some cases, this will have been in response to the cap.^e
- **Couples separating so that the lower benefit cap limit would apply.** The overall effect of this type of behavioural response on welfare spending is ambiguous as the combined value of the two separate claims might be higher or lower than the original single claim. Perhaps unsurprisingly therefore, there was no evidence for this in DWP's evaluation.
- **A more buoyant labour market than we expected** at the time of the original costing (see Chapter 3) with higher employment and lower unemployment, is also likely to have reduced the pool of households eligible for the cap.

We estimate that this caseload component accounts for £150 million (81 per cent) of the overall £185 million difference between the original and revised costing (Chart C).

A lower-than-expected weekly loss for those that were capped

At the time of the costing, uprating was assumed to be in line with CPI inflation, which as we saw in Chapter 3, was higher than we forecast in 2010. The uprating assumption for most working-age benefits was changed at Autumn Statement 2012, when it was limited to 1 per cent for three years from 2013-14. DWP's evaluation suggests that the mean weekly loss for capped households from a November 2013 cohort was around £70, which is around £25 lower than that assumed in the costing. We estimate that this element contributes £69 million (37 per cent) of the overall difference between the original and revised costing (Chart C).

Chart C: Benefit cap: revised 2014-15 costing



Source: DWP, OBR

These two factors more than explain the overall shortfall between original and revised costing, leaving an unexplained residual of £34 million (18 per cent). While our revised estimate of an £80 million saving is produced on a like-for-like basis with the original costing (therefore only capturing the amounts capped) neither is the true reflection of the measure's impacts on the public finances. For example, the true fiscal impact of the cap should also include the taking up of other government support. An additional £110 million of discretionary housing payments (DHP) was made available over 2013-14 and 2014-15 to offset the impact of the benefit cap. Around £55 million was used in relation to the benefit cap.^f DHP is made through DWP's departmental expenditure limits so does not classify as welfare spending, but raises government borrowing in the same way. Finally, it is notable that, while the benefit cap reduces welfare spending by a relatively small amount in aggregate terms, it does so with a relatively large reduction in award for the relatively low number of households that are capped.

^a The cap applied to the main out-of-work benefits at the time, including housing benefit, child benefit, child tax credit and carer's allowance. Its original scorecard name was 'Total household benefit payments capped on the basis of average take-home pay for working households'.

^b At Budget 2012 the exempt groups were expanded to include ESA claimants in the support group. A nine-month grace period for claimants who were working in the preceding year was also introduced. These were contained in the single scorecard line, 'welfare reform bill: amendments', which also contained other measures. The cap was lowered in the July 2015 Budget, effective from November 2016, to £385 a week for couples and lone parents (£442 in London) and £258 for singles (£296 in London). We do not consider the impacts of this later measure.

^c DWP, *Benefit cap impact assessment*, July 2012.

^d DWP, *Benefit Cap: Analysis of outcomes of capped claimants*, December 2014.

^e DWP, *Post-implementation effects of the Benefit Cap*, December 2014.

^f DWP, *Use of Discretionary Housing Payments: financial year 2013/14 and financial year 2014/15*.

Direct behavioural effects in welfare costings

4.20 Estimating the behavioural response to a policy change is often the most uncertain aspect of any costing. This can be particularly true for welfare policies, where the data quality is usually reliable, and the costing models are well established. It is important to make a distinction between *direct* behavioural effects, which are captured within the costing, and wider *indirect*

behavioural effects, which feed into our fiscal forecast via our economy forecast. These are the two processes by which we capture dynamic responses to policy changes.

4.21 The distinction between direct and indirect behavioural effects is often a pragmatic one. If a change in entitlement rules for one benefit causes claimants to move to another, that would be best captured in the costing, as the overall economic effects are muted. If the overall effect of a package of welfare reforms were expected to increase or reduce total employment, it would be better to capture that in our labour market forecast. This way the employment impact can be reflected in *all* the relevant tax and spending forecasts via the economic determinants that underpin those forecasts.

4.22 When quantifiable, direct behavioural effects are captured within the policy costing, increasing or reducing savings or costs relative to the static costing. Examples of direct behavioural responses include:

- **Changes in take-up:** welfare spending is directly affected by the take-up of different benefits – i.e. the proportion of those eligible for a benefit who actually claim it. Evidence suggests take-up increases with the size of the cash entitlement and it seems reasonable to also expect rates to be higher for simpler claim processes or when benefits are more widely promoted (Chart 3.4). Measures that influence these factors will affect take-up and spending. Claiming for one benefit may also prompt a claim for another, through increased awareness and contact with the benefits system. This may be true more widely too – for example, for a period, our forecasts assumed that the rollout of the Dilnot social care reforms (until they were shelved) would result in some individuals realising that they were eligible for attendance allowance.
- **Changes in circumstances:** for many benefits, eligibility or entitlement is linked to specific claimant circumstances – for example, household income, employment status or the number of dependent children (see Chapter 3 case studies). Where policy seeks to change the relationship between claimant circumstances and eligibility or entitlement, claimants may respond by actively seeking to change their circumstances, either to avoid a penalty or to realise a gain. This behaviour often directly affects the costing, but predicting its extent is challenging, as the claimant themselves may be unsure how a change in policy might affect their circumstance.¹⁶
- **Moving onto other benefits:** policies that reduce spending on one benefit may affect the desirability and take-up of other benefits for claimants with multiple entitlements. In these instances, we may expect post-behavioural savings to be lower than the static estimates as claimants substitute to other benefits. In other cases, policy may intentionally seek to widen eligibility for one benefit to draw certain claimants from another. In practice, it is unlikely that all claimants will choose to transfer, and this too affects the costing.

¹⁶ Some behaviours are subject to less uncertainty than others. For example, age-related changes in qualifying criterion can be objectively measured while others, such as ill-health or disability, are more difficult to determine. We would also expect greater behavioural responses to larger and more prominent changes in policy.

- **Changes in compliance behaviour:** a policy that, for example, introduces tighter restrictions on eligibility for some benefits may result in an increase in fraudulent claims or existing claims continuing in error due to lack of awareness, which should be allowed for in the costing. In contrast, some measures have directly sought to improve compliance, for example generating savings by reducing error and fraud rates.

Case study 8: High income child benefit charge

At Spending Review 2010 the then Government announced that it would withdraw child benefit from those households in which at least one adult earned above the higher-rate threshold (HRT) for income tax, from January 2013. At the time of the costing, the HRT was expected to reach just under £43,000 in April 2013. At Budget 2012 the original announcement was superseded by a second measure that made two major changes to the design of the policy:^a

- First, it introduced the high-income child benefit charge (HICBC). Households with at least one person earning over £50,000 received the same amount of child benefit but had to pay an income tax charge equivalent to 1 per cent of their child benefit award for every £100 they earned above £50,000, so that the proportion withdrawn reached 100 per cent from £60,000 upwards. This replaced the initial ‘cliff-edge’ threshold that would have made parents earning just over the HRT worse off than those earning just under it, with a relatively smooth taper for households containing an adult earning between £50,000 and £60,000. That said, the novel approach of seeking to claw back spending issued to households via the benefit system through an individual’s self-assessed (SA) income tax return made this measure far more complex than its predecessor.
- Second, replacing the HRT as the key policy threshold with a flat £50,000 limit made HICBC more generous initially, but with that generosity eroding over time due to fiscal drag (since the HRT was due to rise each year with CPI inflation).

Both costings relied on the Treasury’s inter-governmental tax and benefit model, which is a micro-simulation model populated with survey data from the Family Resources Survey (FRS). This meant there was the usual issue of there being a long lag between the survey data and the measure taking effect, five years in the case of the 2010 costing, reduced to three years for the 2012 version of the costing that we focus on in this case study, which used 2009-10 FRS data.

The costing considered several behavioural responses, including affected households choosing to: not register for child benefit in the first place; opt out of receiving payments if they had already registered; and continue receiving the award before paying it back to HMRC via the HICBC charge. The first two behavioural responses reduce child benefit expenditure, while the third increases tax receipts. The estimates of all three were subject to a high degree of uncertainty, with the latter subject to additional uncertainty around how effectively HMRC would police the charge. The costing also contained a fourth behavioural response, namely households reducing their taxable income to fall below the threshold and maintain their awards.

The amended measure was estimated to generate £1.7 billion for the Exchequer in 2014-15. Our revised estimate is £1.1 billion, a 37 per cent shortfall. The simple explanation is that fewer

households were affected by the measure than originally assumed, although, as we discuss below, there are several other potential factors at play.

Fewer households breached the £50,000 limit than assumed

The original costing assumed that 1.5 million families would exceed the £50,000 threshold, while our latest view is that the actual number was around 1.3 million, 12 per cent lower. The main reasons are:

- As discussed in Chapter 3, average earnings grew more slowly than our 2010 forecasts predicted, and this is likely to be the most important reason why fewer households breached the threshold.
- Sampling issues in the FRS, which typically underestimates the number of higher-income households, might have also played a part.^b
- Possibly more households than expected reduced their taxable income, for example by increasing pension contributions.
- It is also possible that in some households, one partner increased their income while the other decreased it, so that both remained under £50,000 threshold, therefore maintaining their child benefit award. This latter behaviour was not allowed for in the costing.

Chart D shows that fewer households than assumed exceeding the £50,000 threshold contributed £0.3 billion (39 per cent) of the overall shortfall between the original and revised costing.

Other reasons why the number of affected households was lower than expected

As with other costings of new taxes, we needed to consider how HMRC would identify those affected, how it would deal with non-compliance, and how it would reclaim awards through the self-assessment system. There was also uncertainty around households' awareness, especially given the unusual part-tax, part-welfare nature of this measure. For example, in some households it might be that it is the mother that takes responsibility for the child benefit claim but the partner that is breaching the HICBC threshold. In this instance would the latter even be aware that they had the HICBC liability?^c And for many households with otherwise simple income tax affairs, they would have no reason to engage with the self-assessment system if they were unaware of their need to do so as a result of the introduction of this measure.

The outturn data shows that around 300,000 paid the charge in 2014-15 compared to the original estimate of 1.2 million and that around 450,000 chose to opt out of receiving a child benefit payment, compared to the assumed 100,000. Therefore, some of the loss in expected tax receipts is offset by lower-than-assumed welfare spending.

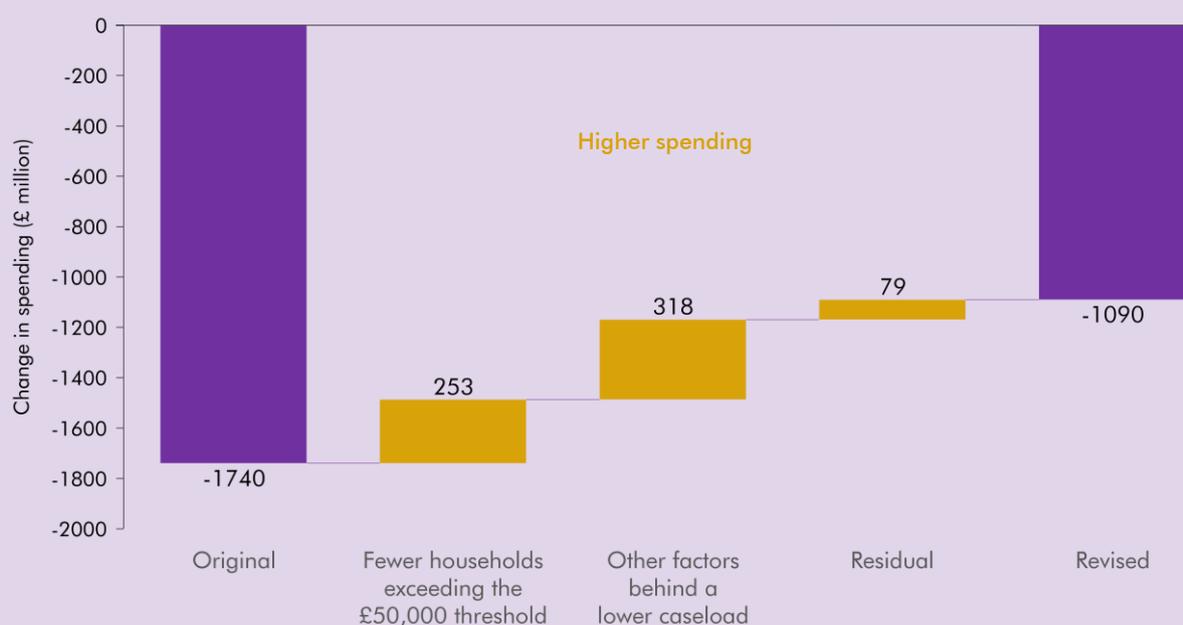
Non-compliance, whether intentional or not, seems to have played a part in the much lower-than-expected number of households paying the charge, though we are unable to isolate its effect from others. HMRC's analysis shows that 80 per cent of those that they contacted as part of compliance checks agreed that they had incurred a 2014-15 HICBC liability. The number of individuals declaring a HICBC liability has remained in the 300,000 to 400,000 range since, raising around £0.4 billion each year, despite the £50,000 threshold being fixed in cash terms.

The number opting-out of child benefit has been rising, reaching 624,000 by 2020 according to the most recent statistics.

We estimate that the remaining shortfall in the caseload – above and beyond that explained by fewer-than-assumed households breaching the £50,000 limit – accounts for a further £0.3 billion (49 per cent) of the overall difference between the original and revised costing (Chart D). That leaves a non-caseload residual difference of £0.1 billion (12 per cent).

It is worth noting that the £50,000 threshold has never changed and under current plans will remain at that level until at least 2027-28, with fiscal drag bringing more households in scope every year. For costings of this kind, where the five-year estimate is not representative of the longer-term impact, we usually highlight the longer-term costing when the measure is first scored, but this permanent cash freeze was not explicitly stated at the time of the measure’s introduction. If the threshold had risen with CPI inflation beyond the original costing – i.e. between 2015-16 and 2027-28, it would have reached £66,500 by 2027-28 (a third higher than £50,000).

Chart D: High income child benefit charge: revised 2014-15 costing



Source: HMRC, OBR

^a The original scorecard titles of the two measures were ‘child benefit: remove from families with a higher rate taxpayer from January 2013’ and ‘child benefit threshold and taper’. The scorecard costing of the first estimated it would reduce spending by £2.5 billion in 2014-15. The costing for the second explicitly included the effect of the first within its pre-measures baseline, which explains why its scorecard costing was to increase spending in 2014-15 by £0.6 billion, since it was relatively more generous in the short term. The combined £1.7 billion estimate is lower than the sum of the two original costings since the former was revised down at Budget 2012.

^b For example, when it is compared to HMRC’s personal tax model, which is populated with data from the Survey of Personal Incomes.

^c HMRC did run a publicity campaign when HICBC was introduced, including writing to 800,000 taxpayers and further correspondence after that.

Indirect behavioural effects and impacts on the labour market

- 4.23 Indirect or ‘second round’ behavioural effects are those with wider-reaching whole economy impacts and can be generated from individual policies or from an overall policy package. Estimating ‘indirect effects’ separately from direct behavioural effects allows us to make a judgement on the net fiscal effect of a policy package, accounting for potential overlaps between measures, possible interactions with monetary policy and the prevailing economic environment.¹⁷
- 4.24 Indirect effects can be incorporated in two main ways. First, we use fiscal multipliers to capture the impact of the entire policy package at a Budget or other fiscal event. These estimate the overall impact of the Government’s announcements on aggregate demand. Second, we can make policy-specific changes to our forecast of the supply side of the economy, including changes in labour market participation.¹⁸

Fiscal multipliers and aggregate demand

- 4.25 For demand side effects we use fiscal multipliers to estimate the combined effect on aggregate demand of all the policy decisions in a fiscal statement. As we set out in our November 2020 EFO *“these multipliers capture the indirect effects of the fiscal measures on activity over and above their immediate effect on demand and through raising private incomes and spending. However, they also take account of the consequent upward pressure this puts on wages and prices and the monetary policy response by the Bank of England necessary to keep inflation at target”*. This latter point explains why our multipliers taper to zero over the five year forecast horizon.
- 4.26 We have a ‘standard’ set of multipliers that vary for different categories of spending, with an AME multiplier of 0.6, reflecting the relatively higher marginal propensity to consume for some benefit recipients, though offset by higher-income retirees who may be more inclined to save a portion of their payment. The AME multiplier sits between a higher multiplier for departmental spending and a lower one for tax cuts. We are not bound by these multipliers, or indeed the assumption that they taper to zero by year five of the forecast and can diverge from them either due to the prevailing economic environment or the emergence of new evidence.¹⁹

The supply side of the economy and potential output

- 4.27 The labour market is perhaps the main channel through which welfare policies may influence the supply side of the economy. Indeed, governments often cite greater labour

¹⁷ An alternative way of considering the macro-level impact of policies is to use a process known as ‘dynamic scoring’. This approach estimates the indirect effects for each measure individually and attributes the full impact on the public finances to that measure’s costing. However, generating dynamic costings in this way is impractical – given the vast number of policies we must scrutinise at a fiscal event – and places a huge additional burden on the amount of detailed evidence that would be required to support the dynamic effects. There is no evidence that such an approach would be more accurate – indeed there is good reason to think a dynamic scoring approach would be less effective at accounting for interactions across measures. For more on this, see our *Briefing Paper No.6: Policy costings and our forecast*.

¹⁸ We are also able to make changes to individual variables, such as changes to our inflation forecast following a change to excise duties, or changes to property transactions if there is a change in stamp duty land tax.

¹⁹ For more, see our December 2019 *Forecast evaluation report* and our November 2020 *EFO*.

market participation as a policy rationale (the June 2010 Budget report states that “reforms will reduce benefit dependency and promote work”).

4.28 Welfare measures can influence the labour supply choices at the individual and household level by changing the relative returns from being in or out of work and, for those already in work, the returns to working more hours. But changes in financial incentives need to be weighed against non-financial impediments, such as family responsibilities, health issues and skills barriers.²⁰

4.29 Supply-side effects, unlike those on the demand side, are assumed to be longer-lasting. When assessing the potential indirect behavioural impacts of welfare policies, we consider several factors, including:

- **The net change in financial incentives:** nine of the ten case studies assessed in this paper restrict the generosity of welfare payments, mainly by imposing tougher conditions on eligibility or entitlement. Lowering the value of *out-of-work* welfare payments relative to employment income, whether directly or indirectly, creates a financial incentive to enter employment, and the reverse would be true for policies that increase the relative value of welfare payments. Restrictions to *in-work* benefits reduce the return from moving into work so reduce the incentive to work. Increasing the rate at which in-work benefits are withdrawn as incomes rise reduces the incentive to work more hours, since in effect it increases the marginal tax rate on the earnings from an additional hour worked. Financial incentives tend to be more effective when they are more transparent and more widely promoted.²¹
- **The magnitude of the change in incentives:** some households may choose not to seek to change their behaviour until the gain or loss is sufficiently large. Even then, their ability to do so is likely to be linked to the prospects for the wider economy, including the number and types of vacancies. Aggregate behavioural responses to changes in financial incentives are estimated using ‘elasticities’ but these are often highly uncertain, including whether there are thresholds below which the elasticity is zero.
- **The materiality of the behavioural effect:** in isolation, most welfare policies are too small to affect our labour market forecast materially, so our assessment tends to focus on larger measures or on the aggregate impact of the policy package announced at a fiscal event. This has been important, since policies such as the freeze on most working-age benefits have applied across in-work and out-of-work benefits, which has meant that the aggregate effect of elements that improve the incentive to work have often been offset by those that worsen it. And it has often been necessary to consider those changes alongside changes to the income tax personal allowance.

²⁰ For more on how government policy might affect potential output see our *Briefing Paper No.8: Forecasting potential output – the supply side of the economy*.

²¹ As part of our scrutiny of policy costings we do routinely ask for information on how the Government will promote new measures. A good example is the introduction of tax free childcare in 2017. Similarly, if a new measure is getting widespread publicity in the media, such as offering free TV licences for those claiming pension credit, then we take that on board when estimating the level of take-up.

- **The durability of a policy change:** a policy that leads to a long-term change is likely to be more impactful than a shorter-term measure in terms of its underlying impact. For example, the raising of the State Pension age permanently increases the numbers of years that people work, boosting long-run labour supply. Temporary measures are more likely to have greater impact on the timing of activity – for example, bringing it forward into a period where its treatment is temporarily more generous – rather than the long-run level of activity.
- **The additionality of a policy change:** governments routinely intervene to try and increase labour market participation and, therefore, the impacts of these ongoing interventions are implicitly captured in our pre-measures labour market forecast. For us to include an indirect effect above and beyond that already included in the baseline, a new measure must be demonstrably different and additional to previous ones. Additionality also requires that other government policy, including outside welfare, does not offset the effects of a given measure.
- **The availability and quality of supporting evidence:** the Coalition Government introduced over 40 welfare measures in 2010 including some large-scale reforms such as the introduction of universal credit and changes to disability benefits. The speed and scale of these changes was unprecedented in recent history, so past evidence on the effects on incentives was scarce. Occasionally we are presented with the results from a small-scale trial or pilot preceding a full policy rollout, but these are often not representative of wider implementation and may therefore be of limited use – for example, because those taking part in the trial do so voluntarily or because those staffing the trial are more expert or are given more time per intervention than would be the case once the policy was rolled out to all DWP job centres.

Case study 9: Income support: extend lone parent conditionality to those with children aged five and above

This June 2010 Budget measure, phased in between May and November 2012, meant that lone parents with a youngest child aged 5 or 6 were no longer entitled to claim income support (IS) solely based on being a lone parent. Instead, if they were deemed able to work, they would have to apply for jobseeker's allowance (JSA), which, among other things, required claimants to be actively seeking work. Those lone parents that were unable to work would be migrated to either employment and support allowance (ESA) or other health-related benefits.

The policy was the fourth phase of the 'lone parent obligations' (LPO) policy, the tightening of out-of-work support for lone-parents that began in November 2008. The LPO initially applied to lone parents with a youngest child aged 12 and above, before the threshold was progressively lowered, to age 10 and above from October 2009 and to age 7 and above from October 2010.

The costing was informed by the initial results of the earlier LPO phases and assumed that 10 per cent of affected IS claimants would move off benefits, with 6 per cent moving into work. Of the remaining 90 per cent, most were expected to move onto JSA with a smaller proportion moving to ESA. The more challenging conditionality around JSA was also expected to generate savings by slowing the inflow of new claimants, relative to the pre-measure position.

Our revised estimate is that the measure reduced welfare spending by £275 million in 2014-15, just over 50 per cent *higher* than the original £180 million costing. The primary reason is that the reduction in the caseload was more pronounced than initially expected, with this only marginally offset by a lower-than-assumed rise in the value of awards.

The population of 5- and 6-year olds was larger-than-expected

The general population of 5- and 6-year olds grew by more than assumed in the pre-measures forecast, so it is likely that we underestimated the number of cases would have been affected by the measure. We estimate that this element explains £6 million (6 per cent) of the overall £95 million difference in between the original and revised costing.

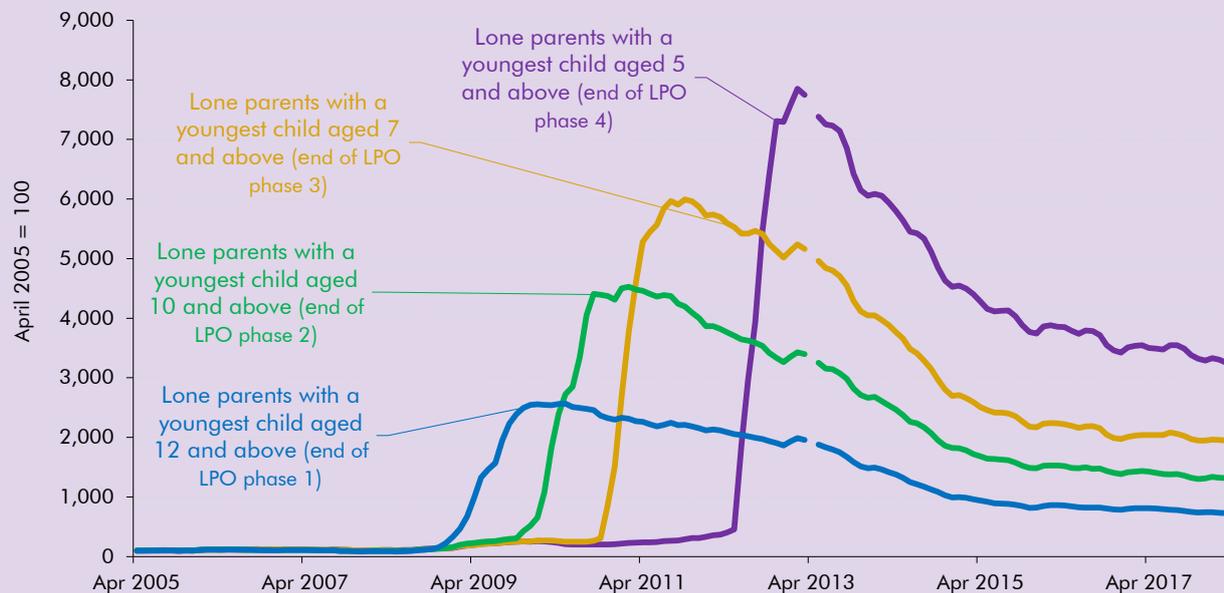
It is likely that more people moved off benefits than assumed in the costing

DWP's evaluation of the first three phases of the LPO, published in 2013, suggested that these policies may have prompted significant labour market responses. This included an increase of 8 and 10 percentage points in the share of affected lone parents who were in work 12 months after the introduction of the first and second phase respectively.^a The 6 per cent assumption in this costing reflected the likelihood that lone parents with younger children faced greater barriers to returning to work, as shown in the wider employment data.^b But the evaluation also noted that parents with younger children might have more recent experience of work and that this might encourage relatively more of them to move into work.

The number of lone-parent IS claimants with a youngest child aged 5 and 6 declined by around 85,000 during phase 4, with the equivalent JSA caseload increasing by close to 50,000. Indeed, over the four phases of the LPO the number of lone parents claiming JSA increased by a peak of around 150,000 cases. Chart E shows that the number of JSA cases followed a similar pattern across all four phases of the LPO, with a rapid initial rise followed by a decline over time, in part

due to the improving labour market conditions discussed in Chapter 3. It might also indicate that the costing under-estimated the extent to which lone parents would exit JSA over time.

Chart E: Number of lone parents claiming JSA across each LPO phase



Source: DWP

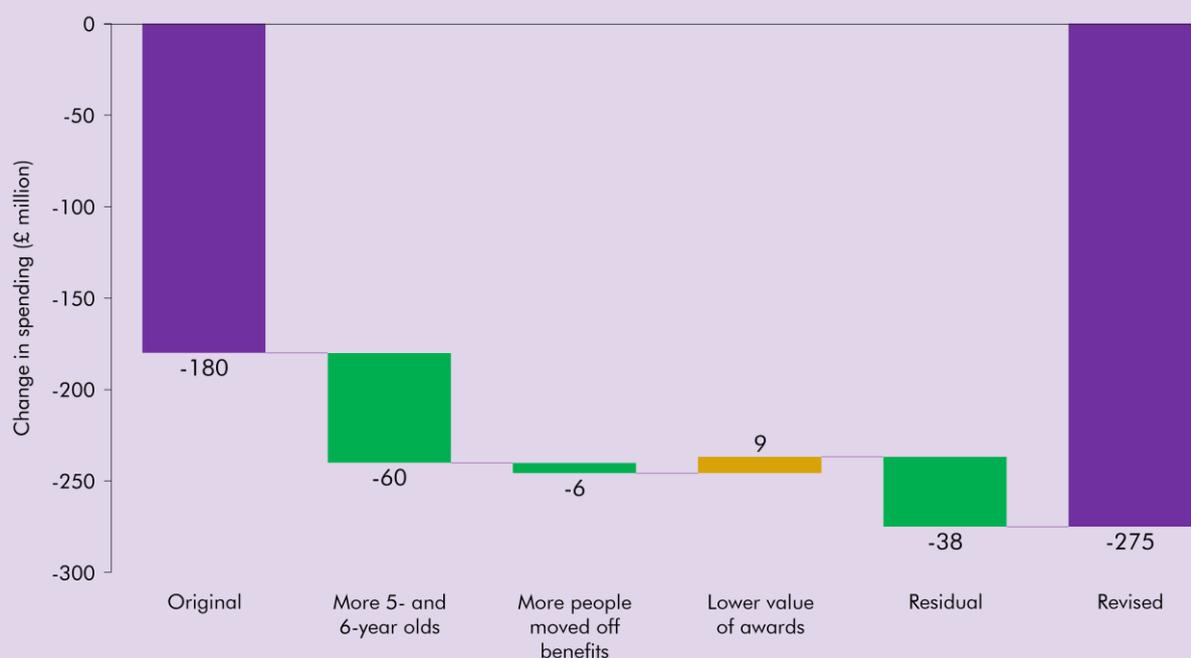
At the time, we adjusted our forecast of claimant count unemployment to reflect that some lone parents were expected to move from IS to JSA. In one sense, this was an 'indirect effect' since the policy led to a behavioural response that was factored into our economy forecast, but in reality, it simply aligned our claimant count forecast with the static costing of the measure since the claimant count was an administrative measure of unemployment. We did not alter our overall unemployment forecast, so in effect assumed that this measure changed the ratio of the administrative count to the broader survey-based measure of unemployment. DWP's evaluation did not estimate any displacement effects, so it is not clear the extent to which the increased employment among lone parents added to total employment in the economy, rather than displacing other potential workers. This wider effect is what we aim to capture in our economy forecast when considering the true indirect effects of a policy.

Overall, it does seem likely that the difference between the revised and original estimates can be partially explained by larger-than-expected numbers of lone parents moving into employment instead of other out-of-work benefits, including some doing so with a lag. We estimate that this element accounts for £60 million of the overall difference in the costing (63 per cent, Chart F).

Lower value of awards

The per-case cost of remaining on IS or other out-of-work benefits was slightly lower than assumed, reducing the savings from the measure relative to expectations. One of the factors behind this will be the Autumn Statement 2012 decision to limit uprating for working-age benefits to 1 per cent for three years from 2013-14.^c We estimate that lower-than-expected awards reduces the saving by £9 million, leaving an unexplained residual of £38 million.

Chart F: Income support: extend lone parent conditionality to those with children aged five and above: revised 2014-15 costing



Source: DWP, OBR

^a DWP, *Lone Parent Obligations: an impact assessment*, July 2013.

^b ONS' *Labour Force Survey* shows that lone parent employment rates rise with the age of the youngest child.

^c As discussed in Chapter 3, CPI inflation outstripped expectations, making the limit on uprating more impactful. The uprating measure was scored at Autumn Statement 2012.

Case study 10: Housing benefit: removal of the spare room subsidy

This June 2010 Budget measure, more commonly known as the 'bedroom tax', reduced the level of housing benefit (HB) payment for social-housing households who "occupy a larger property than their family size warrants".^a If a property in the social-rented sector was deemed to be under-occupied – meaning it had one or more spare bedrooms – then the household's HB entitlement was capped at the level of the average regional rent for a property "appropriate to their household size and structure" (aligning the treatment with that for private-sector tenants). The policy was amended at Budget 2012, replacing the average regional rent cap with a percentage reduction in HB entitlement and increasing the age limit, from 60 to the qualifying age for pension credit.^b It came into effect in April 2013.

The original 2010 costing used 2008-09 data from the Family Resources Survey to: identify households that would be affected by the new rules; establish average social-sector rents; and estimate the pre-measure level of HB entitlement. The estimate of post-measure HB entitlement additionally used information on household size (from private-sector tenants) before applying the new rent caps. The static costing was, in effect, the difference in entitlement multiplied by the number of affected households.

The costing allowed for several potential behavioural responses, including households downsizing to smaller accommodation, compensating for the loss in income by moving into work (or increasing their hours), and letting out (or otherwise occupying) the unoccupied room(s). Most of these assumptions were based on limited evidence.

The 2010 costing predicted savings of £490 million in 2014-15, with an additional £60 million (mostly from claimants aged 60 and above) coming from the Budget 2012 amendments. Our revised estimate, based on administrative data, is that this measure reduced HB spending by £365 million in 2014-15, a third lower than the combined £550 million original estimates. The difference is largely due to fewer-than-expected households being affected by the measure. The effects of two other factors – a higher-than-expected reduction in households' HB awards and some modelling updates that took place between the 2010 and 2012 costings – largely offset.

Fewer households were affected than assumed in the costing

The combined original costings estimated that 660,000 households would be affected by the measure.^c In the event, the number of affected households peaked at close to 550,000 immediately after the measure took effect, and fell to an average of around 470,000 during 2014-15. Chart G shows that the lower-than-expected caseload accounts for £173 million (94 per cent) of the £185 million overall difference between the original and revised costing.

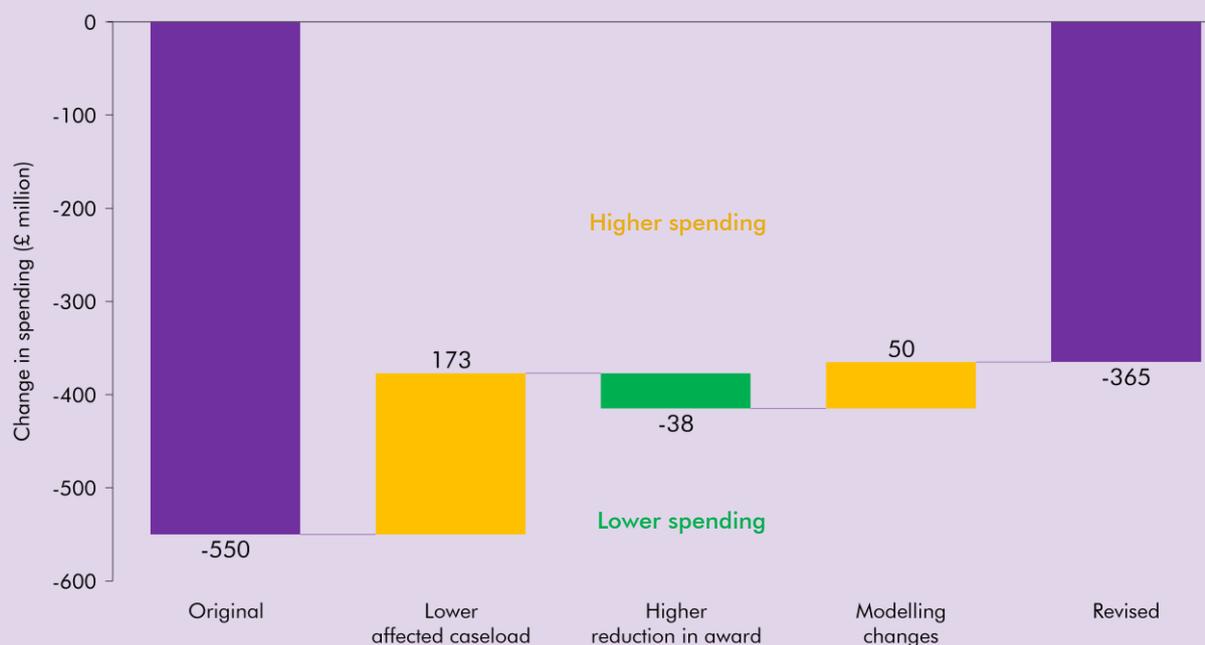
This caseload shortfall is set against the backdrop of a larger overall working-age HB social-sector caseload than we forecast in June 2010, suggesting that behavioural responses might have played a larger role than we thought at the time. The key behavioural effect in the costing was that 'under-occupied' households might seek to maintain their HB awards by bringing in tenants. This effect brought the number of households expected to lose from the measure down by a third in the original costing, though this was highly uncertain.

In the event, DWP's 2015 evaluation suggests that other behaviours played a larger role than we expected, with households responding in a variety of ways including finding work, increasing hours worked, and relocating, including moving into the private-rented sector.^d It also notes that 17 per cent of households initially affected by the measure ceased to be so by 2014.^e

The reduction in weekly housing benefit exceeded the original estimate, but was offset by modelling changes

The combined original costing assumed that affected households would, on average, lose £14 of HB a week. The administrative data show the actual loss was slightly higher at £15 a week, which increases savings from the measure by £38 million (21 per cent) (Chart G). This was more than offset by modelling changes made when costing the Budget 2012 amendment, including using 2009-10 FRS data, which lowered the saving by around £50 million (Chart G).

Chart G: Housing benefit: removal of the spare room subsidy: revised 2014-15 costing



Source: DWP, OBR

As with the benefit cap, this measure led to additional government spending through the discretionary housing payment (DHP) scheme. DHP is made through DWP’s departmental expenditure limits so does not classify as welfare spending, but raises government borrowing in the same way. This measure received a DHP allocation of £60 million in 2014-15, following on from a £55 million allocation the year before. Actual costs outstripped these allocations in both years, with DHP spending against this measure amounting to just over £100 million in 2014-15 and a two-year total of around £180 million.^f These additional payments do not strictly count against this measure but clearly the savings from the costing needs to be seen in the context of higher public expenditure from elsewhere.

Finally, as with the findings from the benefit cap, this measure achieves a relatively modest reduction in overall welfare spending, but does so by substantially lowering benefits for a relatively small number of affected households.

^a Its original scorecard name was ‘social sector: limit working age entitlements to reflect size of family from 2013-14’ and it has also been referred to as the ‘social-sector size criteria’ and the ‘under-occupancy penalty’.

^b This had no impact on female claimants, whose pension credit qualifying age was 60, but brought more male claimants into scope since the qualifying age for them was 65. The Budget 2012 measure was captured within the ‘welfare reform bill: amendments’ measure on the Treasury’s scorecard.

^c DWP, *Housing benefit: under occupation of social housing impact assessment*, June 2012.

^d DWP, *Evaluation of removal of the spare room subsidy*, December 2015.

^e Some households were able to avoid the measure temporarily by exploiting a loophole, but this was closed from April 2014 and therefore only has a negligible impact on our revised 2014-15 estimate. The loophole enabled households to escape deductions for under-occupancy during 2013-14 following an unintentional widening of transitional protection that was put in place for private tenants that year. The loophole affected tenants that had received housing benefit continually (and for the same property) since 1 January 1996. Though it was closed in April 2014, households could still claim back any HB deducted during 2013-14. DWP’s evaluation suggests that this affected relatively few households.

^f DWP, *Use of Discretionary Housing Payments: financial year 2013/14 and financial year 2014/15*.

5 Conclusions and lessons learned

5.1 This paper has examined how the range of welfare policy measures announced over the past 13 years have performed relative to our initial costings, paying particular attention to 10 case studies taken from policy announcements made in 2010 itself. The exercise helps to illustrate the challenges we face when scrutinising a costing, demonstrates how that scrutiny has evolved in line with our accumulated experience, and highlights lessons for the scrutiny of the fiscal implications of future welfare policy changes. In this final chapter we:

- review the various factors that can drive deviations between the initial estimates and final cost of welfare policy changes;
- discuss how our approach to scrutinising costs has evolved; and
- summarises the lessons for scrutinising future welfare policy costings.

Explaining deviations in welfare policy costings

5.2 The ultimate costing of welfare policy changes can deviate from their original estimates for a variety of reasons, many of which have been identified in this paper and illustrated by our 10 case studies. These include:

- **Differences that arise from the pre-measures forecasts** that underpin every costing. These baseline forecasts are subject to considerable uncertainty and any differences between forecast and the eventual outturn data will, all else equal, have a knock-on effect to the costing. Since we take the pre-measures economic and fiscal forecasts as an exogenous input into the costing, this is an inherent uncertainty, one that is detached from the costings scrutiny process. Significant differences between baseline economy forecasts and outturn for employment, earnings growth and inflation in the aftermath of the late-2000s recession explain much of the difference between the original and revised costings for several of our case studies.
- **Major policy decisions being made without an accompanying OBR forecast.** The measures that were announced and scored in Spending Review 2010 did not come alongside an OBR forecast and, as we have explained, several of these costings were changed considerably just a month later when we published our November 2010 *EFO*. When significant policy decisions are made outside a forecast their announced costings are inconsistent with our current outlook and unlikely to be subject to the same level of scrutiny (though we will report our own estimates in subsequent *EFOs*).

- **Subsequent policy changes.** Our forecasts include the costings based on the policy that is announced by the Government at the time, but this is not necessarily the same as the one that is eventually implemented. Policy details often change, and these can have a major effect on the costing, though we do reflect these changes in subsequent forecasts. Our case studies include multiple examples where the policy has been amended before implementation, with these amendments typically lowering the expected level of savings from the measure. Similarly, measures that reduce welfare spending might lead to the introduction of measures that increase spending elsewhere, such as discretionary housing payments. These findings echo those from our December 2019 *Welfare trends report*, when we re-estimated the package of welfare spending reductions announced in the July 2015 Budget, where a series of post-announcement concessions and policy reversals also led to welfare spending being reduced by significantly less-than-expected. We have regularly shown that governments may find it harder sticking with pre-announced medium-term spending reductions when the eventual implementation of those plans will generate cash losers from one year to the next. This has parallels with our observation that governments have tended to increase DEL spending when departmental allocations are set in Spending Reviews, compared to their previous medium-term spending assumptions.¹ By contrast, measures that generate savings through a change in threshold, such as freezes in the uprating of benefits, seem to be more credible commitments, arguably because they are more politically deliverable, as the cash amount received by a claimant remains unchanged.
- **Uncertainties relating to the data and modelling.** The costing models used by DWP and HMRC, and the data that underpin them, are generally reliable representations of the tax credits and benefits systems. But use of both departments' main models is hampered by the long lag until full administrative or survey data become available. This introduces uncertainty when projecting forward from the base year to the year that a policy will take effect, albeit that both departments use statistical techniques to mitigate against this and improve their models' representativeness. The lag in data can be particularly challenging when, as with our ESA case study, the introduction of new benefits is subject to policy amendments before even survey data, let alone outturn data, is available.
- **The difficulties associated with assessing operational delivery.** We have repeatedly found that costings that rely on changes to operational implementation or departmental systems have delivered benefits more slowly than originally estimated. These costings require us to scrutinise a wider set of factors than more traditional costings, so we now routinely ask questions about issues like staff recruitment, staff training and IT infrastructure. We assess project timetables, and the amount of contingency built into them, and compare these to the performance of similar previous measures. Similarly, there are welfare policy costings that rely, to some degree, on carrying out claimant's health assessments. As we point out in our ESA case study, these assessments do not always go according to plan, something we have also

¹ Atkins, G. and L. Lanskey, *The OBR's forecasting performance*, Working Paper No. 19, August 2023.

reported on in relation to personal independence payment and universal credit.² Past welfare measures where operational delivery has failed to meet initial targets have almost always led to higher-than-expected spending.

- **Estimating the direct and indirect behavioural responses.** Many welfare policies announced since 2010 have affected claimants' behaviour, but understanding these in a sufficiently detailed way that can be reflected in individual costings remains a challenge. This challenge is amplified when a very large package of measures are announced simultaneously, as they were in both 2010 and 2015. For example, some of the case studies in this paper provide a better understanding of the ways in which claimants can respond to a policy change and how such behaviour might affect welfare spending and the wider economy, particularly the labour market. However, until we can reliably isolate and quantify such effects, making judgements around behaviour in welfare costings will remain highly uncertain. Departments do sometimes evaluate their policies, but this remains relatively rare and where it happens is not usually undertaken sufficiently long after introduction to determine the full impact, and rarely has a focus on the public spending consequences.

How our scrutiny of welfare policy costings has evolved

5.3 To try to address some of these challenges, we introduced a number of changes to the way in which we scrutinise policy costings over the past 13 years. These include:

- **A dedicated policy costings team within the OBR.** At the time of the June 2010 Budget, the OBR was still an 'interim' entity with fewer than ten permanent staff. That number had increased slightly by the time of the November 2010 forecast but did not include any senior staff working full-time on policy costings. As the OBR has expanded, so too has our attention on policy costings, which now has its own team that supports the Budget Responsibility Committee in the scrutiny of costings and jointly co-ordinates the costings process with the Treasury and other departments.
- **Higher quality information and more regular engagement on costings.** The knowledge gained by those organisations involved in producing policy costings has led to a general improvement both in the quality of costing notes and the sharing of information. There are also mutual benefits to be gained by working with departments on a year-round basis, for example, we may be able to scrutinise changes to costing models outside of the busiest periods or provide earlier advice on costing methodology. More regular engagement also helps flatten the intensity of workload peaks just before a fiscal policy statement, fostering improved scrutiny and mitigating the potential for errors.
- **Increased transparency around policy costings with more coverage in our publications.** As well as scrutinising the costings of new measures, we also monitor, evaluate and

² *Welfare trends report*, January 2018 (for universal credit) and *Welfare trends report*, January 2019 (for personal independence payment).

report on costings after they have been announced. Recostings allow us to update our forecasts to reflect the latest information but also improve our evidence base when we are asked to scrutinise costings similar to those that we have seen before. Policy costings now has a full chapter in the *EFO* as well as a host of supplementary online material. This includes a *Supplementary scorecard* that provides a more detailed breakdown than its Treasury counterpart, a *Policy measures database* that contains a similar breakdown for all costings back to 2010, a *Policy costings uncertainty ratings database* that lists our subjective assessment of the uncertainties for every costing dating back to 2014 and a *Policy risks database* that discusses a host of policies that are, at the time of publishing our forecast, insufficiently firm to include.

- 5.4 These developments allow us to swiftly adapt to events occurring outside our forecast schedule, such as with the costings database we published during the early months of the pandemic. We initially populated the database with our own estimates and updated it to reflect new announcements and emerging data. This ensured that the fiscal impact of the Government's interventions was communicated publicly, helping it play a more prominent role in the discourse. This adaptability proved invaluable during the unprecedented events of autumn 2022. The disruption caused by factors beyond our control necessitated the unusually rapid delivery of a forecast, complete with an extensive package of new policies.

Lessons for the scrutiny of future welfare costings

- 5.5 We are committed to enhancing the quality of our policy costings scrutiny, and to that end we finish by collating the main lessons that we have learned while scrutinising welfare policy costings, encompassing both those discussed in this paper and more generally.
- 1 **Costings should utilise the timeliest available data sources, and ideally administrative data where available.** Where possible the administrative data should drill down as close as possible to the affected population. If a timely source is unavailable, then the data lags should be dealt with by calibrating costings models or through other means. Similarly, if a costing is based on survey data, then calibration may be required if there is a low sample size for the affected population.
 - 2 **When a new benefit is introduced, departments should concurrently design appropriate administrative data collection and surveys, building in better functionality at an earlier stage.** But we should remain wary about the reliability of new data sources, which can sometimes take a while before they reach a steady state.
 - 3 **Estimating the change in the level of take-up following a policy change can be challenging, but taking on board lessons from previous measures boosts the evidence base when making those judgements.** Take-up tends to be higher for higher cash awards. Policies that expand the coverage or encourage take-up of a given benefit can also prompt claims for other benefits as contact with the welfare system increases. Policies that reduce spending on one benefit may affect the desirability and take-up of another benefit for claimants with multiple entitlements. The uncertainties around each

- can be amplified by wider economic uncertainty, for example during the cost-of-living crisis and the pandemic.
- 4 **We need to ensure that departments can deliver policies in the way that is set out in the costing note.** This is one of the key lessons we have learned since 2010 and encompasses both major reforms like the introductions of universal credit and personal independence payment and countless smaller-scale measures. Policy reforms need to be accompanied by a credible operational delivery timetable that is sufficiently resourced and based on evidence from previous interventions, with suitable contingencies included.
 - 5 **Policy costings need to account for interactions with other contemporaneous changes within the benefits and tax systems, with the overall fiscal impact adjusted accordingly.** The scale of the welfare spending reductions announced in 2010 were unprecedented, so exercises such as this are valuable in understanding these interactions. There are times when restrictions around the sharing of Budget-sensitive information means that we are unable to ask departments to cost the impact of another department's policies on their spending. For example, this meant that the March 2023 childcare measures did not include impacts on welfare spending that fed through DWP and HMRC models, relying instead on broad-brush estimates of our own. We subsequently asked DWP and HMRC to provide revised estimates to include in our next forecast, highlighting the issue in both accompanying *EFOs*. We routinely ask, but cannot insist, that the Treasury shares such information with affected departments.
 - 6 **Welfare policy changes can have both temporary effects on aggregate demand and effects on aggregate supply that persist into the medium term.** We account for these separately in our forecasts. For the former, we will continue to monitor the available evidence and take on board the full range of views around fiscal multipliers. For the latter, we have committed additional resources to improving our understanding of how fiscal policy changes can impact the supply side of the economy. This includes the empirical evidence around how changes in welfare policy affects labour market participation, weighing up factors such as the impact of financial incentives and non-financial barriers. To further support our knowledge base we have recently expanded our advisory panel of economic and fiscal experts.
 - 7 **Realtime monitoring of data on fraud and error (F&E) can help improve the reliability of future compliance measures.** The fiscal consequences of compliance measures tend to be uncertain, and this has been compounded recently by uncertainties around the baseline forecast, following a spike in F&E during the pandemic. Measures that change entitlement rules can also drive changes in the levels of F&E, which need to be estimated in the costing.
 - 8 **Departments should undertake more regular ex-post evaluations of their policy costings, to improve the quality of future exercises.** There is a compelling case for more policy evaluation, and for this to be built into the policy design framework from the outset, such as the provision for data collection. We will continue to ask departments to

provide us with recostings of previously announced measures at each forecast and to report on their findings. Revisiting past costings yields highly valuable insights.

- 9 **Cooperation on policy costing models outside fiscal events can help ensure timely and reliable costings when policies are announced or changed.** While policy costings should utilise current OBR forecasts where possible, ensuring the underlying assumptions are consistent with our latest view, working together with the Treasury and other departments outside fiscal events can improve the costing's methodology and mitigate against uncertainties. Given the imperative to develop policies rapidly in response to recent shocks, this early investment in model development pays dividends.
- 10 **Maintaining a high degree of transparency around policy costings increases the level of public understanding, and helps compensate for the sometimes-limited information provided by the Government.** This also invites greater external challenge, which we will use to further reinforce our scrutiny of policy costings.

