

2. MEASURING ETHNIC PARITY

2.1 Definition of 'Ethnic Parity'

A natural definition for there to be *ethnic parity* in outcomes from Jobcentre Plus services and programmes is if there is no difference, on average, in the outcomes for an ethnic minority participant compared to an 'otherwise identical' white participant. Where parity does not exist, there will be either an ethnic penalty – if ethnic minority customers experience worse outcomes than otherwise identical white customers – or an ethnic premium – if ethnic minority customers experience more favourable outcomes than otherwise identical white customers.

This definition is an ideal one, and the aim of this report is to provide measures of the degree of parity or the extent of the penalty (premium) that most closely approximate the ideal (or 'true') ones.¹

A first objective of the project was to investigate the current methodology used by the Department for monitoring Jobcentre Plus performance for ethnic minorities and if there were any ways the measures could be improved. The following sections 2.2 and 2.3 thus briefly outline the current approach to assess ethnic parity and highlight some of its problems. Section 2.4 will then propose ways to address these shortcomings and describe a new approach, while section 2.5 will outline the methodology.

2.2 The current approach to measuring ethnic parity

To monitor the performance of employment programmes – New Deal for Young People, New Deal 25plus and New Deal for Lone Parents – the Department currently uses a monthly measure based on the difference in the proportions of Ethnic Minority and White programme leavers who accessed jobs:

$$\frac{\# \text{ Ethnic Minority New Deal leavers into jobs} / \# \text{ Ethnic Minority New Deal leavers}}{\# \text{ White New Deal leavers into jobs} / \# \text{ White New Deal leavers}}$$

The current approach for measuring overall parity in Jobcentre Plus performance uses an extract of data from the Labour Market System (LMS) to show the number of customers gaining jobs in a particular quarter as a proportion of all customers with *any* LMS activity (e.g. job/opportunity referrals or starts, interviews, adjudications or sanctions) recorded in the same quarter:²

$$\frac{\# \text{ Ethnic Minorities into jobs} / \# \text{ Ethnic Minorities with any LMS activity}}{\# \text{ Whites into jobs} / \# \text{ Whites with any LMS activity}}$$

Note that ethnic parity is not currently assessed separately for individuals making benefit claims (other than as part of LMS activity in the overall Jobcentre Plus measure); this project has however sought to estimate ethnic parity for individuals on Income Support (IS), Incapacity Benefit (IB) and Jobseeker's Allowance (JSA) as well.

¹ Further discussion of interpretational issues of this definition can be found in Chapter 4, section 3.6.

² An overall parity measure is calculated, together with measures for selected groups of Wards and Local Authority Districts with high ethnic minority and unemployment concentration.

2.3 Problems with the current approach

There are a number of methodological problems with the approach outlined above. These are summarised in the box, and discussed in some detail in the corresponding subsections.

The parity measure currently used by the Department:

1. simply compares outcomes for Whites and Ethnic Minorities, without making any attempt to compare 'otherwise identical' individuals;
2. considers
 - a. all individuals who leave the New Deal during a given period, thus selecting on the outflow (New Deal measure);
 - b. all individuals with any recorded interaction with Jobcentre Plus during a given quarter, thus confounding stock and flows (Jobcentre Plus measure);
3. is based on the number of activities and not on the number of individuals;
4. is focussed on job entry alone, de facto treating it as an absorbing state.

2.3.1 Selection bias

Attributing all the observed difference in outcomes between Whites and Ethnic Minorities to their ethnicity ignores the possibility that these two groups may differ in terms of characteristics other than ethnicity that also affect the outcomes of interest.

In other words, the simple difference in the observed outcomes for White and Ethnic Minorities groups would provide a biased estimate of the true ethnic parity if there were systematic differences between the two groups that also affect outcomes. Such discrepancy in observed outcomes may arise because of differences in characteristics that can potentially be observed by the analyst (observables) as well as in characteristics that are not observed by the analyst (unobservables). Of course it is the data that the analyst has access to which determines what is observed and what remains unobserved. In this data available for this research one can observe, say, gender and previous labour market history, but not innate ability or motivation.

For example, Bangladeshi women have lower employment probabilities than many other ethnic groups and one might think this is explained by discrimination on the grounds of ethnicity. However, these women also have much lower levels of qualification and are more likely to experience language difficulties³, which, if taken into account, might be expected to significantly reduce the correlation between ethnicity and employment outcome – the observed ethnic 'penalty'.

³ See for instance, Owen, D., Green, A., Pitcher, J. and Maguire, M. (2000), "Minority ethnic participation and achievements in education, training and the labour market", *DfEE*, Research Report No. 225.

The current DWP measure, obtained by comparing the proportion of ethnic minority customers who obtain a job with the proportion of white customers who obtain a job, would only be valid under the special – and unlikely – case in which the distribution of all observable and non-observable background characteristics which determine job outcomes are the same for all ethnic groups accessing Jobcentre Plus services or a given programme.

2.3.2 *Sampling frame*

To monitor the performance of the New Deal programmes, the Department currently considers the difference in the proportion of recorded job-starts between certain dates for White and Ethnic Minority participants who are observed to *leave* the New Deal. For this measure, sample selection is thus based on *outflow*, which raises (at least) two issues.

First, such a measure does not consider those customers who did not record an exit. It thus misses the important group of individuals who simply continue to remain on benefits, and evidence suggests that these individuals are more likely to be from ethnic minority groups. Consider the following example, where the current measure reported that 80% of the ethnic minority participants who leave the programme exit to a job and only 60% of whites, but where only 5% of ethnic minority customers leave the pool over that time period, while 50% of white customers do.

The second issue has to do with the selection process out of unemployment. Since the current measure selects the sample based on an outcome, i.e. leaving the programme, selection into the group of leavers is likely to be non-random and there may be systematic differences between the way that Whites and Ethnic Minorities have been selected into this group. In particular, if ethnicity affects exit rates from unemployment, white and ethnic minority individuals who are observed to leave unemployment will differ in terms of unobservables. To see why, consider that the highest quality (in terms of labour market) individuals tend to leave benefit/unemployment first; if rates of outflow differ by ethnicity, then the quality of the outflow of claimants will differ by ethnicity, which could lead to biased results.⁴

A similar reasoning applies to the measure used by the Department to assess Jobcentre Plus overall. The sampling frame currently used is all individuals with a recorded interaction with Jobcentre Plus during a given quarter, which involves sampling from both the inflow and the stock. To fix ideas, consider the case of a programme. Although this sampling frame allows one to capture the *inflow* into the programme during a particular quarter, a large proportion of the individuals being analysed will have entered prior to this and will have been on the programme for a long time (the *stock*). Again, it is quite possible that selection into this group of customers with long durations is non-random, in that the stock of Whites and the stock of Ethnic Minorities who are observed to still remain on the programme in a given quarter may differ along both observed and unobserved dimensions, depending on the possibly

⁴ Since individuals of higher ability (irrespective of ethnicity) leave unemployment faster, if *ceteris paribus* Whites also leave unemployment faster, then white customers who are observed to leave will be on average of lower ability than Ethnic Minorities who leave.

different process by which Whites and Ethnic Minorities leave the programme.⁵ Again, this could lead to biased results.

2.3.3 *Spells versus individuals*

The existing approach is based on spells, not on individuals. This means that people who had more than one spell on a programme (or benefit) in any given time period (in the case of the existing approach, a quarter) will be counted in the analysis according to the number of spells they had on a programme/benefit in that quarter. Thus repeated job exits by the same individual within the quarter of interest will count as corresponding 'successes'. However, to exhibit repeated job exits, the individual has necessarily to have come back onto benefits in this quarter, an indicator of lack of success in that job. This lack of success is not captured in an analysis of spells.

This approach not only ignores the issue of sustainability in employment, but in fact it *rewards* low sustainability. This is particularly serious if different ethnic groups vary in the extent to which they 'cycle' on and off programmes/benefits. As an extreme case, consider a situation in which most of the white customers cycled back and forth between short jobs and benefits, while ethnic minority customers mostly waited for a good job match and hence kept the job. In this scenario, the current measure would show a large ethnic penalty.

2.3.4 *Outcome measures*

The existing approach only considers moves into jobs as an outcome. Only looking at exits into jobs explicitly ignores possible future spells and in particular the potential for return to benefits.

Furthermore, focusing exclusively on exits into jobs does not allow one to get a full picture of any differences in the quality of labour market destinations. By contrast, differences in job retention and employment probability over time would allow one to gauge potential differences in the sustainability of jobs that white and ethnic minority customers go into.

Similarly, looking at new outcome measures relating to whether individuals continue with benefit receipt is likely to be of considerable policy interest.

2.3.5 *Ethnicity not recorded*

The existing approach is only calculated for customers who have their ethnicity recorded in the administrative data, and thus omits all customers whose ethnicity is recorded as: 'none selected', 'prefer not to say' or 'no personal contact'. From both a methodological and policy perspective, it would be useful to consider how many of these individuals there are and whether they represent a distinct group in themselves, with systematic differences in their exit patterns (when compared to Whites) in terms of the outcome measures considered.

⁵ Since individuals of higher ability (irrespective of ethnicity) leave unemployment faster, if *ceteris paribus* Whites also leave unemployment faster, then Whites who are among the stock of unemployed will be on average of lower ability than Ethnic Minorities who are still unemployed.

2.4 Proposed new approach to measuring ethnic parity

This section sets out the details of the new proposed approach to measuring ethnic parity, highlighting how the shortcomings described in the previous section have been addressed.

The primary aim of this project is to calculate the degree of parity or the extent of penalty/premium by seeking to compare 'otherwise identical' individuals. The ensuing requirements are set out in the box, and discussed in some detail in the corresponding subsections.

In order to compare 'otherwise identical' individuals, the suggested methodology strives to:

1. carefully control for observed differences between Ethnic Minority and White customers using appropriate methods;
2. reduce the likelihood that individuals differ in unobserved dimensions as much as possible via a suitable choice of sampling frame;
3. choose individuals as the unit of analysis, not spells;
4. obtain a fuller picture by considering a number of carefully constructed outcomes.

2.4.1 Selection bias

Interest lies in the extent to which there is *ethnic parity* of outcomes from Jobcentre Plus services and programmes, by considering whether there are any differences in the outcomes for ethnic minority participants when compared to 'otherwise identical' white participants.

As discussed in Section 2.3.1, the current measure is obtained by simply comparing the proportion of ethnic minority customers who obtain a job with the proportion of white customers who obtain a job; it thus ignores the possibility that white and ethnic minority customers differ, on average, more than just in terms of ethnicity.

To prevent parity measures being affected by such 'selection' bias, one needs to control for any differences between ethnic groups in observed and unobserved characteristics that may affect their outcomes, such as the individual's background and labour market history. By doing this one will be able to gain a better understanding of how much the observed ethnic difference in labour market outcomes, such as job entry rates, is due to differences in the characteristics of White and Ethnic Minority groups, and how much can be attributed solely to ethnicity, that is, to the 'ethnic penalty'.

Two main methods are available to control for *observed* differences between individuals belonging to different ethnic groups:

- standard regression techniques (ordinary least squares – OLS): and

- matching (in particular, propensity score matching) methods.

It is important to underline that both types of method only allow the researcher to control for observable differences between individuals, that is characteristics that are measured and recorded in the data (see section 4.3.1 for caveats and section 3.3 for a review of the available variables). Furthermore, matching and regression differ in the way in which they control for such observable differences.

Under suitable assumptions, the difference-in-differences method by contrast allows for *unobserved* differences between Whites and Ethnic Minorities that are constant over time and that affect their labour market outcomes in a constant way. This approach has also been used for part of the analyses, mainly as a sensitivity check.

The various methodological approaches are discussed in some more detail in Section 2.5.

2.4.2 Sampling frame

Section 2.3.2 has raised some serious concern that the sampling frames underlying the current measures might differentially miss important groups of individuals (in particular, non-leavers), as well as introduce bias in the composition of Whites and Ethnic Minorities considered.

These concerns do not relate to differential selection into the programme, but rather to differential *selection into the analysis sample*, and that this differential selection might be driven by unobservables. It is therefore proposed to analyse benefit and programme *inflows*; if one focuses on the inflow, selection into the programme is, by definition, selection into the analysis sample.

To be operational, one needs to decide on the details of the inflow window, in particular about (a) its length, and (b) its starting point. There are pros and cons for each of these choices.

A larger inflow window yields a larger sample size, thus increasing the robustness of the models as well as the probability of being able to produce more subgroup analyses. However, a larger inflow window limits the time period over which to measure labour market history and to assess outcomes (for the latter in particular one needs a sufficiently long evaluation horizon to be able to assess sustainability of employment). Finally, a larger window might run the risk of ‘straddling’ periods where there were significant changes to benefit/programme eligibility and/or alterations to procedures.

The choice of the starting point of the inflow window relatively far back in time allows a longer horizon over which the outcomes can be assessed, which is particularly important to assess sustainability of employment. On the other hand, it would impact on the extent to which the analysis can be seen as up-to-date, as well as limit the period of availability for data on labour market history.

Based on current data availability, reliable information on previous labour market history is available from June 1999, while individual employment outcomes can be evaluated until December 2004.

In order to obtain a balance between an analysis that is as up-to-date as possible and data that are rich enough for the task at hand, it was thus agreed that the appropriate choice of inflow window would cover the 2003 calendar year.

Crucially this allows the analysis:

- to consider the existing versions of the New Deal programmes, and to focus on IS and IB recipients who have had a Work Focused Interview⁶,
- to have a period of at least three years to measure previous labour market history, and
- to follow entrants at the end of 2003 for a whole year, which would allow most of them to participate fully in the programmes and also have the opportunity to record a period of sustained employment, as discussed below.

Table 2.1 Distribution of completed durations of programme or benefit spell by ethnicity, 2003 inflow

	NDYP		ND25Plus		NDLP	
<i>Duration (months)</i>	<i>Ethnic Minority</i>	<i>White</i>	<i>Ethnic Minority</i>	<i>White</i>	<i>Ethnic Minority</i>	<i>White</i>
3	41.0	42.5	28.7	29.1	27.1	38.6
6	62.1	62.4	50.9	55.1	47.9	62.6
9	73.9	74.8	66.2	71.4	65.6	77.4
12	82.6	83.4	81.0	84.0	79.0	87.9
15	89.7	90.1	89.8	91.9	86.9	93.4
18	94.2	94.4	94.3	95.6	92.1	96.1
	JSA		IS		IB	
<i>Duration (months)</i>	<i>Ethnic Minority</i>	<i>White</i>	<i>Ethnic Minority</i>	<i>White</i>	<i>Ethnic Minority</i>	<i>White</i>
3	43.0	58.8	9.1	12.7	4.3	6.3
6	66.2	78.6	27.1	31.7	22.9	27.7
9	81.8	89.7	41.1	46.9	39.7	44.7
12	88.9	93.8	48.1	54.6	49.9	54.2
15	92.8	96.1	53.3	59.8	55.5	59.5
18	94.9	97.1	57.5	63.6	58.8	62.8

Note: Each figure denotes the percentage of the programme/benefit inflow that has left the programme/benefit within x months from inflow, x = 3, 6, 9, 12 15, 18. Duration is until recorded first exit.

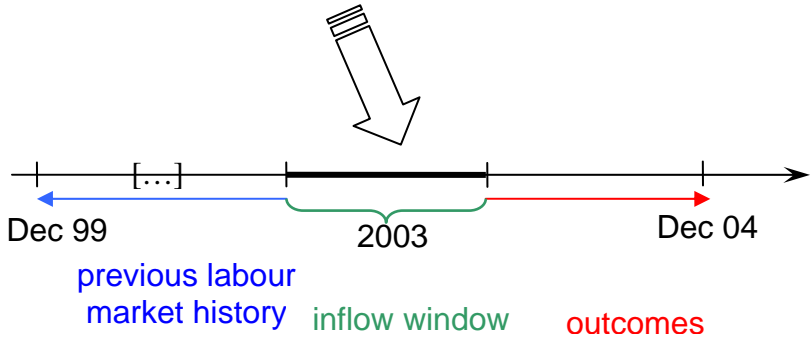
⁶ Work Focused Interviews (WFI) were introduced in 2003. As motivated in more detail in Section 4.1, attention is restricted to benefit recipients who have had a WFI. On substantive grounds, the WFI represents at least some type of 'treatment'; on practical grounds, ethnicity is much better recorded for customers who have had a WFI.

Table 2.1 shows the distribution of completed durations on the three New Deal programmes and for the three benefit groups (IS, IB and JSA), separately for ethnic minority and white customers who started the programme or benefit during 2003. Specifically, the table relates to the duration of the qualifying spell in the relevant analysis sample (see Chapter 3 for details of how analysis samples are defined). It shows the proportion of individuals whose qualifying spell lasts no more than three months, no more than six months, etc.

Table 2.1 shows that for JSA and the three New Deal programmes, between 80 and 90 per cent of qualifying spells last no more than 12 months (the time over which benefit and employment outcomes can be observed). The rate of exit from IS and IB is much slower, with between 45 and 55 per cent of qualifying spells lasting 12 months or less. Durations in the raw data are slightly shorter for white customers than ethnic minorities across all programmes/benefits.

To obtain some measure of job quality in terms of sustainability⁷ within a 12-month window, individuals would need to have started work by the beginning of month 10 after entry. Table 2.1 shows that between 65 and 75 per cent of individuals on a New Deal programme, 80 to 90 per cent of JSA claimants, and 40 to 45 per cent of IS and IB claimants have left the relevant programme/benefit 9 months after inflow.⁸ Considering a 12-month window thus seems to provide the right balance.

Figure 1: Outline of the approach



The chosen timeline is highlighted in Figure 1:

- individuals being analysed are the inflow into a programme or benefit during 2003;
- their past labour market history is tracked back to three years before inflow;
- their labour market outcomes are evaluated over a one-year period.⁹ In particular, a person who entered a programme or benefit on the last available day

⁷ Note that the Department defines an individual as being in sustained employment if they have been continuously employed for at least 3 months.

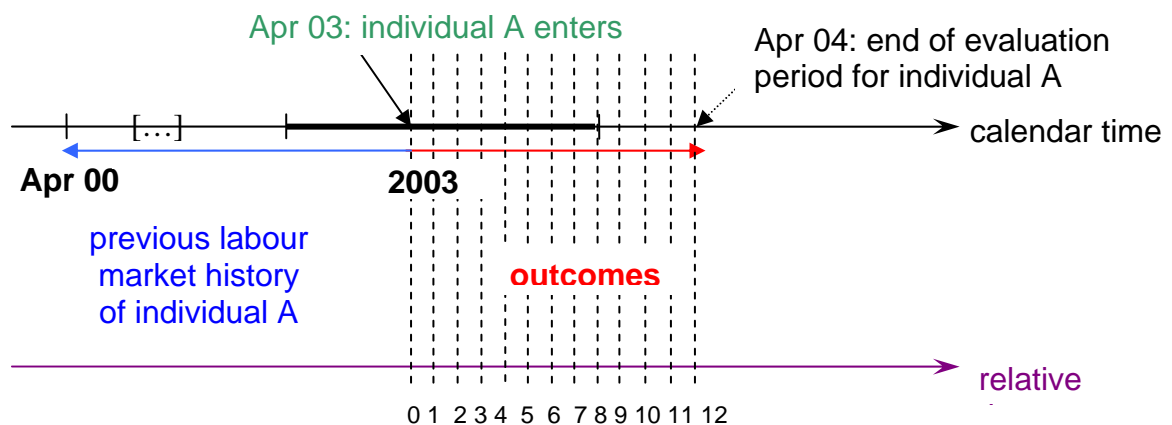
⁸ Of course, this means that for individuals who enter employment in months 10, 11 or 12 after inflow, sustainability of employment cannot be assessed within the 12-month window.

⁹ Note thus that even for individuals for whom one could potentially observe outcomes for up to two years (those who entered a programme/started claiming a benefit at the beginning of January 2003), only outcomes for the first 12 months after entry are considered.

of the inflow window, 31st December 2003, would have 12 months to move through the programme/benefit and record some type of employment or benefit-related outcome.

Figure 2 further outlines the approach by sketching an individual example. Individual A enters the programme in April 2003; at the moment of inflow, relative time for that individual is set to zero. More generally, relative time is set to zero at each individual's recorded programme or benefit entry date. The ethnic parity measures are then constructed by recording post-inflow outcomes month-by-month, i.e. after 1, 2, ... and 12 months from the individual's inflow (note that 30-day periods were used for this purpose rather than calendar months). For the current example, this corresponds to May 2003 to April 2004 (Section 2.4.4 below further discusses the chosen outcome measures). An individual's history is constructed for the three years prior to inflow, which in this case amounts to recording individual A's labour market status from April 2000 to April 2003.

Figure 2: Individual A enters the programme in 2003



2.4.3 Spells versus individuals

The key question this report wants to answer for each ethnic minority group relates to how different their labour market outcomes would have been if they had been white. An individual-level approach is thus adopted, which samples individuals starting a programme/benefit during the inflow window and then records their subsequent outcomes.

In contrast to the spell-based approach previously used by the Department, the suggested method is able to capture (differential) recidivism and employment sustainability, providing in particular a more accurate picture of those people who cycle on and off benefit. Thus an individual who starts drawing benefits several times in the inflow window will be recorded once and his/her outcome will be his/her status at a specific point in time. The spells-based approach would count this person several times and may possibly record several job exits, whilst using our approach the individual will be given the same weight as any other person and their exit to jobs would only count as a positive outcome if it were sustained (rather than being followed by a return to benefit).

2.4.4 Outcomes

Section 2.3.4 has identified and discussed a number of shortcomings in the way outcomes are measured by the current approach. By focusing on first exit only and indeed on exit to employment only, this approach ignores any ethnic differences in return to benefits ('recidivism') and in the quality (sustainability) of labour market destinations.

In order to capture (differential) recidivism and employment sustainability, outcome measures that follow individuals over time will be considered, with labour market status assessed month-by-month, up to one year after inflow. This should allow enough time for customers to find jobs as well as for them to display recidivism, and should thus give an accurate picture of the medium and more long-term extent of ethnic parity.

Furthermore, labour market status will be measured in different dimensions:

1. employment probability,
2. probability of being in sustained employment, and
3. probability of being on benefits.

This new way of measuring outcomes thus looks beyond exit into jobs and considers whether customers are in employment or on benefit over time (specifically, monthly) following inflow into the programme/benefit (cf. Section 2.4.2). Specifically:

1. To allow one to obtain a fuller picture of any differences in the quality of destinations, the new approach assesses differences in job retention or employment probability over time.
2. An additional measure of employment sustainability is also used to directly assess any potential differences in the sustainability of jobs white and ethnic minority customers go into.
3. Continued benefit receipt is analysed in order to capture both the extent to which individuals who are non-employed remain on benefits and the extent to which individuals who are employed still collect benefits (in particular, income support).¹⁰

To measure the percentage of individuals who are employed (or claiming benefits) x months since programme/benefit start (where x goes from 1 to 12 and months are measured as 30-day periods), the following three options have been considered:

- a. employed (or on benefit) for 15 or more days during month x ;
- b. employed (or on benefit) at any time during month x ; and
- c. employed (or on benefit) at multiples of 30 days since inflow, i.e. on the $(30 \cdot x)^{\text{th}}$ day since inflow.

Option (a) has been chosen after separately testing these alternatives and finding that they did not make any significant difference in terms of the resulting ethnic parity

¹⁰ Note that the proposed measure of benefit receipt will not be able to separately identify these two effects, nor will it be able to check that individuals in employment are claiming IS (rather than something else).

measure.¹¹ The reason is that, in any given month, the vast majority of individuals are either employed/on benefit for the whole month or for none of the month – so all three measures give the same answer.

As a last implementation detail to measure benefit dependency, note that the constructed indicator takes account of both active and inactive benefits¹², although it is not able to distinguish between the two. In particular, a move from an inactive (e.g. IB) to an active (e.g. JSA) benefit – which may be regarded as a positive outcome by the Department – would not be recorded.

Finally, sustained employment was measured from month 3 to 12 after inflow and required the individual to have been *continuously* employed for at least 3 months up to then.¹³ Thus from month 3 onwards, in addition to testing whether there are any ethnic differences in the probability of being in employment, the analyses will also test whether there are any ethnic differences in the probability of being in continuous employment for the previous 3 months.

2.4.5 Ethnicity not recorded

This report treats customers whose ethnicity is not recorded as a separate ‘ethnic’ group. The extent of ethnic parity is thus estimated for the various ‘ethnicity unknown’ groups of each benefit and programme type.

2.5 Methodology

2.5.1 Overview

The main analysis will use propensity score matching techniques (‘matching’) to measure ethnic parity in Jobcentre Plus programmes and mainstream services. As explained in more detail below, these methods are more flexible, in the sense that they impose fewer restrictions on the data, than standard regression-based methods. On the other hand, the latter methods are easier and faster to implement, and at times are found to produce very similar results to the more complex and time-consuming matching estimators. Ethnic parity has thus also been measured using regression-based techniques, and comparisons between the two methods were pursued to assess the reliability of the regression-based methods for the case under analysis.

¹¹ Focusing on specific dates as done by option (c) would miss short employment or benefit spells that fall between these key dates, a shortcoming not suffered by the other two options, which measure outcomes over a period of time. Option (b) did not seem demanding enough.

¹² Specifically, JSA, compensation whilst on a New Deal option (achieved by including NDYP and ND25plus spells as benefit spells), Basic Skills, WBLA, IS or IB all counted towards the measure of benefit receipt. Note that this measure is defined by considering whether the individual is on benefit on day 1, day 2 . . . day 30 of the month of interest; thus if an individual is on more than one type of benefit on a particular day, it is only counted as a single day on benefit.

¹³ So to count as sustainably employed e.g. in month 3, an individual would need to have been *continuously* employed for *all* 30 days of month 1 (i.e. the month following inflow date), month 2 and month 3. This is a slightly stricter definition from the one for being employed in a given month, which only requires the individual to have been employed for at least 15 of the 30 days).

Duration modelling (outlined in section 2.5.4) adds an important time dimension to the analyses, by ascertaining whether particular ethnic groups are able to find employment more quickly than others.

While matching, regression and duration methods only control for observed differences between Ethnic Minority and White customers, difference-in-differences methods (outlined in section 2.5.5) provide evidence on the robustness of the findings to the presence of uncontrolled (i.e. unobserved) individual differences between ethnic groups.

2.5.2 *Measuring ethnic parity through matching methods*

The main research strategy of this analysis involves using propensity score matching techniques (or ‘matching’) to address the key question that needs answering for each ethnic minority group: How different would their labour market outcomes have been if they had been white?

To construct such a counterfactual for an ethnic minority group, one needs to select, from the pool of white Jobcentre Plus customers, a comparison group who ‘looks the same’, in terms of observed characteristics X , to the customers of the ethnic group under analysis. Matching is the best available method for selecting such a matched (or re-weighted) white comparison group in which the distribution of individual and local area characteristics that might affect labour market outcomes, X , is as similar as possible to the distribution of X in the ethnic minority customer group of interest.

More specifically, matching allows one to match every customer from a particular ethnic minority group to a similar white customer (or to a group of white customers). In essence, the matching procedure re-weights Whites so that they look as close as possible to the relevant ethnic minority group of interest in terms of factors (X) that might affect labour market outcomes.¹⁴ To ensure comparability, ethnic minority customers for whom no suitable white comparator can be found (i.e. who fall outside of the so-called common support) are excluded from the analysis (this issue is taken up again in Section 4.3.2). The estimate of the ethnic penalty/parity/premium is then obtained by comparing the mean labour market outcomes of the ethnic minority group with the mean outcomes of the appropriately matched/re-weighted white comparison group. This can be done for each programme or service and for each ethnic minority group of interest.

Matching techniques are thus able to control in a flexible way for the effect of observed background characteristics on labour market outcomes so as to accurately isolate the impact of ethnicity, that is, our measure of ethnic parity.

¹⁴ To implement propensity score matching, a probit or logit regression model can be run, where the dependent variable is equal to 1 for the relevant ethnic minority group and to 0 for the white comparison group. From this model one can estimate the ‘propensity score’, that is, the predicted probability of being from the particular ethnic minority group of interest given the person’s individual and local area characteristics (X). Each member of the ethnic group of interest can then be matched to either the closest white individual or to a group of ‘similar’ white individuals. There are numerous matching methods that can be used, see Blundell, Dearden and Sianesi (2004) for full details.

Note that matching methods – as well as regression methods discussed below – are based on the assumption that one can observe all the outcome-relevant differences between white and ethnic minority Jobcentre Plus customers. Any unobserved differences other than ethnicity between the groups being compared will show up as ethnic penalty or premium (depending on whether the remaining difference concerns an unobserved trait that is unfavourable or favourable to the labour market). The success and reliability of ethnic parity estimates based on either of these approaches will thus crucially depend on the amount and quality of the characteristics we can observe (this discussion is taken up again in section 4.3.1).

2.5.3 *Measuring ethnic parity through regression-based methods*

Compared to matching, standard regression techniques (OLS) have the advantage of being fast and simple to implement, but they will not necessarily overcome biases in estimates of ethnic parity due to the following:

- They may estimate the ethnic penalty by comparing non-comparable individuals using extrapolation (the common support problem);
- They may not weight comparable individuals correctly;
- They typically assume that the ‘ethnic penalty’ is the same for all individuals, preventing it from varying according to individual observed characteristics.

Matching techniques are more flexible and less restrictive than regression models in that they do not suffer from these problems; matching has thus been the primary focus for the research in this report.

On the other hand, regression techniques offer the following advantages compared to matching methods.

First, by imposing a (linear) structure, OLS allows one to obtain more precise estimates (in fact, OLS is the most efficient among the linear and unbiased estimators). This feature is likely to be particularly attractive when disaggregating the analysis and thus reducing sample sizes.

Secondly, the restrictions typically imposed in regression models can be significantly weakened. In particular, in the regression model one can allow the ethnic penalty to vary according to each observable characteristic. Previous research has indeed shown that such a fully interacted regression model can often produce results that are very similar to the ones obtained by matching (Blundell, Dearden and Sianesi, 2005).

Consequently, an important part of the research project has been to assess the reliability of regression techniques by comparing their results to those produced by matching.

2.5.4 *Measuring ethnic parity through duration modelling*

Duration modelling can add an important time dimension to the analysis by ascertaining whether particular ethnic groups are able to find employment more quickly than others – both as documented in the data and once holding observed characteristics

constant. Specifically, such models allow one to assess the potentially different impact of time on the probability that Ethnic Minority and (comparable) White customers leave unemployment for a job. In this set-up, there would thus be 'ethnic parity' if an ethnic minority customer experiences on average the same probability of leaving unemployment over time as does an 'otherwise identical' white customer.

Methodologically, this report implements duration modelling in the form of *survival functions*. Consider a pool of individuals who are unemployed at inflow; the survival function in unemployment at time x is defined as the probability that an individual 'survives' in unemployment for at least x days.¹⁵ In terms of estimation, and in line with the main type of analysis, the aim is to be less parametric as possible, that is, to avoid imposing unnecessary functional form and distributional assumptions. Survival functions have thus been estimated fully non-parametrically (using the Kaplan-Meier estimator), both on the raw samples and on the matched samples. The survival function on the raw samples is a purely descriptive device that allows one to assess whether Ethnic Minority and White customers leave unemployment at the same rate. By contrast, the survival function estimated on the matched samples controls for observable differences in background characteristics between Ethnic Minority and White customers based on the same rationale of matching methods (cf. section 2.5.2).¹⁶

Note that customers who are already employed at inflow have conventionally been given an unemployment duration of one day to prevent them from being dropped from the duration analysis. Specifically, one should see an initial step in the survival function around time of inflow, and indeed a differential step if Ethnic Minority and White customers differ in the extent to which they are already employed at inflow.

A number of formal procedures are available to test whether the survivor functions of the two groups are statistically different from one another.¹⁷ The test of the equality of the two survivor functions implemented in this report is the Peto-Peto-Prentice test. This version is appropriate when the survival functions are thought to vary in ways other than proportionality; it is further unaffected by differences in censoring patterns between the two ethnic groups.

Another useful methodological aspect is that the area under the survival function represents the expected total unemployment duration from inflow to first exit. Hence, the area between the survival function of the Ethnic Minority group and the one of (matched) White group gives the difference in the expected unemployment duration of the two groups.

The discussion so far has considered survival in unemployment (strictly speaking, in non-employment) as the outcome. As outlined in section 2.4.4, however, non-employment and benefit receipt do not necessarily coincide; therefore time on benefits has been additionally looked at via estimation of a corresponding survival function.

¹⁵ In symbols: $S(x) \equiv \text{Prob}\{U \geq x\}$, where U is total unemployment duration since inflow.

¹⁶ For further consistency and comparability with the main matching analysis, survival functions have been estimated on samples matched using kernel matching and after having imposed common support.

¹⁷ These tests are based on non-parametric rank tests for comparing two (or more) distributions.

Before concluding this section it is important to stress that the duration analysis focuses exclusively on the *first* exit from the programme/benefit, thus ignoring any differential subsequent performance of Ethnic minorities and Whites. It has already been discussed at length (cf. sections 2.3.3, 2.3.4, 2.4.3 and 2.4.4) that this might be a potentially serious shortcoming: only looking at first exits explicitly ignores possible future spells, and in particular any ethnic differences in return to benefits ('recidivism') and in the quality (sustainability) of labour market destinations.

By contrast, month-by-month differences in employment probability are able to capture (differential) recidivism and employment sustainability.

As an extreme example, consider a situation in which

- most of the white customers left the programme quite soon, but then returned to benefits very quickly, while
- ethnic minority customers spent on average more time on the programme waiting for a good job match which subsequently ensured their long-term employment.

In this scenario, a duration model would show a large ethnic penalty in terms of first exit from the programme. By contrast, the more 'forward-looking' approach which looks at differential employment rates over the whole observation period would offer a more complete picture: an initial penalty in terms of employment probability (during the time when Whites are leaving the programme faster than Ethnic minorities) which would however quickly disappear and indeed turn into a premium (when Ethnic minorities enjoy sustainable employment while Whites are back on benefits).

The number of transitions of this type that the available observation period of 365 days would allow one to observe is, *a priori*, not clear. If most programme participants remain on the programme or otherwise unemployed for longer periods (a year or more), the duration analysis will not miss too many important patterns; by contrast, if frequent movement in and out of unemployment and benefits is the norm, this analysis would tell only part of the story.

With such interpretational caveats in mind, a minor set of analyses using duration modelling has been carried out for selected comparisons. The aim of these additional analysis was mainly to explore whether the 'story' that emerges from the month-by-month ethnic parity estimates would greatly change in a duration framework.

2.5.5 Measuring ethnic parity through difference-in-differences methods

The idea underlying the difference-in-differences approach is that one can eliminate observed or unobserved differences between Ethnic Minority and White customers that are *constant* over time by taking the differences over time within the groups and subtract the differences between the groups. In this application the difference-in-differences estimator thus measures the excess outcome growth before and after inflow for Ethnic Minorities compared to Whites.

Compared to the approaches outlined above, this method allows for time-invariant unobserved differences between Ethnic Minorities and Whites, in particular it re-

moves differences in unobserved characteristics that are constant over time and that affect individual employment and benefit outcomes in a constant way.

This strategy relies on the assumption that had the Ethnic Minority customers been white, they would have experienced the same *variation* in average employment (and benefit) outcomes as the average outcome variation actually experienced by White customers. This condition requires in particular that had they been white, the Ethnic Minority group would have been affected in the same way by macro economic conditions as the White group, and, for New Deal Ethnic Minority customers, that they would have had the same programme effect as the White customers.

This assumption is thus most plausible when the two groups are very similar, so that asymmetric changes over time in the labour market, programme effects or individual behaviour are less likely. Given however the well-documented differences in important characteristics between ethnic minority groups, this assumption is not likely to be met.

A way to at least partially address this issue is to control for *observed* compositional differences between Ethnic Minority and White customers that are likely to affect their employment dynamics. This conditional difference-in-differences approach has been implemented in a rather flexible regression framework, allowing the ethnic penalty to depend on observed individual characteristics, but restricting the effects of such characteristics on outcome growth to be linear.

A further implementation choice relates to how to deal with multiple pre-inflow time periods. Given that the literature to date has not provided any established solution, the analysis has been carried out in two ways. In the first variant, growth has been measured in respect to a 12-month moving window to capture seasonal effects, i.e. taking the difference between labour market status in the k^{th} post-inflow month and labour market status 12 months before (which is by construction a pre-inflow month given that the analysis post-inflow window is of 12 months).¹⁸ In the second variant, the change in employment (benefit) status at a given post-inflow month has been calculated with respect to the average employment (benefit) probability over the 12-month period before inflow.

Note that to implement an approach which models *changes* in outcomes, individuals with missing or otherwise incomplete 12-month pre-inflow histories have to be excluded from the analysis. Specifically, Ethnic Minority and White customers who have only *first* appeared in the WPLS data during the 12 months before starting the programme/benefit have been dropped.

While the difference-in-differences analysis can provide an important sensitivity or robustness check for the results based on matching methods, there are a few conceptual caveats that have to be borne in mind. These are discussed in section 4.3.3.

¹⁸ Given that the outcome, e.g. employed or not in a given month, is a binary variable, this implies running a regression on a dependant variable which takes on only three values (-1, 0 and 1). While it would thus not be appropriate to treat it as an approximately continuous variable, suitable ways to estimate the standard errors have been used.