Appendix 6.3: Demographic characteristics and commentary on certain SSE and RWE analysis

Introduction

1. The purpose of this appendix is to set out:

   (a) further analysis carried out since publication of our provisional findings, using results from the CMA survey, on the relationship between certain measures of engagement and demographic characteristics; and

   (b) our understanding and views on the analysis and arguments included in:

      (i) the subsection titled 'The PFs fail to identify the main drivers of customer engagement' in section 1 of the report produced by the Authorised Advisers of SSE plc ('SSE') during the disclosure room held after the publication of our provisional findings[1] (the 'post PFs Disclosure Room') (the 'SSE Report');[2] and

      (ii) the report produced by the Authorised Advisers of RWE Npower plc (RWE) during the post-PFs Disclosure Room (the 'RWE Report').[3]

Relationship between engagement and demographic

2. Since publication of our provisional findings, using CMA survey data we explored further the relationships between demographic characteristics of respondents and various measures of engagement. In particular, we looked at results:

   (a) for the following measures of engagement:

      (i) whether a respondent had ever switched supplier;

      (ii) whether a respondent had switched in the last three years; and

      (iii) whether a respondent had switched internally; and

   (b) for the following demographics characteristics, singularly and in combination:

---

[1] This disclosure room was held from 13 July 2015 until 31 July 2015.
[2] In this appendix, we also refer to comments made by SSE at hearings with the CMA.
[3] We have also reviewed the raw analysis conducted by parties within the disclosure room upon which the SSE Report and the RWE Report are based.

(i) annual household income of less than £18,000;

(ii) disability; and

(iii) aged over 65.

3. Taking household income as an example, we extracted results on the percentage of the population:

(a) with household income of less that £18,000;

(b) who had, for example, never switched supplier;

(c) who had, for example, never switched supplier who also had household income of less than £18,000; and

(d) who had household income of less than £18,000 who also had, for example, never switched supplier.

4. These results tell us:

(a) of those who are disengaged, by some measure, what proportion of them have, for example, household income of less than £18,000;

(b) of those with household income of less than £18,000 what percentage of them are disengaged, by some measure; and

(c) how these percentages compare with those in the population as a whole.

5. The results are summarised in Table 1.

Table 1: CMA survey results on the relationship between engagement and demographic characteristics

	Never considered switching	**Switched in last 3 years**	**Switched internally**
1a. Overlap between annual household income of less than £18,000 and engagement 23% of the population have annual household income of less than £18,000			
% of the population who [insert measure of engagement, for example 'never considered switching']	34	64	72
% of those who have [insert measure of engagement] with h'hold income of less than £18,000	30	24	24
% of those with h'hold income of less than £18,000 who have [insert measure of engagement]	44	80	76

	Never considered switching	Switched in last 3 years	Switched internally
1b. Overlap between annual household income of less than £18,000 and engagement – excluding prepayment Excluding prepayment customers - 18% of the population have annual household income of less than £18,000			
% of the population who [insert measure of engagement]	32	74	69
% of those who have [insert measure of engagement] with h'hold income of less than £18,000	24	20	19
% of those with h'hold income of less than £18,000 who have [insert measure of engagement]	42	80	72
2a. Overlap between disability and engagement 12% of the population have a disability			
% of the population who [insert measure of engagement]	34	64	72
% of those who have [insert measure of engagement] with a disability	15	13	12
% of those with a disability who have [insert measure of engagement]	43	81	75
2b. Overlap between disability and engagement – excluding prepayment 13% of the population have a disability			
% of the population who [insert measure of engagement]	32	74	69
% of those who have [insert measure of engagement] with a disability	16	13	13
% of those with a disability who have [insert measure of engagement]	40	80	73
3a. Overlap between those who have annual household income of less than £18,000 and/or have a disability and engagement 31% of the population have annual household income of less than £18,000 and/or a disability			
% of the population who [insert measure of engagement]	34	75	72
% of those who have [insert measure of engagement] with h'hold income of less than £18,000 and/or a disability	39	33	33
% of those with h'hold income of less than £18,000 and/or a disability who have [insert measure of engagement]	43	81	77

	Never considered switching	Switched in last 3 years	Switched internally
3b. Overlap between those who have annual household income of less than £18,000 and/or have a disability and engagement – excluding prepayment Excluding prepayment, 26% of the population have annual household income of less than £18,000 and/or a disability			
% of the population who [insert measure of engagement]	32	74	69
% of those who have [insert measure of engagement] with h'hold income of less than £18,000 and/or a disability	33	28	27
% of those with h'hold income of less than £18,000 and/or a disability who have [insert measure of engagement]	41	80	72
4a. Overlap between those who have annual household income of less than £18,000 and have a disability and engagement 7% of the population have annual household income of less than £18,000 and a disability			
% of the population who [insert measure of engagement]	34	75	72
% of those who have [insert measure of engagement] with h'hold income of less than £18,000 and a disability	9	7	7
% of those with h'hold income of less than £18,000 and a disability who have [insert measure of engagement]	44	79	78
4b. Overlap between those who have annual household income of less than £18,000 and have a disability and engagement – excluding prepayment customers Excluding prepayment, 5% of the population have annual household income of less than £18,000 and a disability			
% of the population who [insert measure of engagement]	32	74	69
% of those who have [insert measure of engagement] with h'hold income of less than £18,000 and a disability	7	5	5
% of those with h'hold income of less than £18,000 and a disability who have [insert measure of engagement]	43	79	73
5a. Overlap between those who have annual household income of less than £18,000 and/or have a disability and/or aged over 65, and engagement 48% of the population have annual household income of less than £18,000 and/or a disability and/or are aged over 65			
% of the population who [insert measure of engagement]	34	75	72
% of those who have [insert measure of engagement] with h'hold income of less than £18,000 and/or a disability and/or are aged over 65	59	52	49
% of those with h'hold income of less than £18,000 and/or a disability and/or are aged over 65 who have [insert measure of engagement]	42	81	74

	Never considered switching	Switched in last 3 years	Switched internally
5b. Overlap between those who have annual household income of less than £18,000 and/or have a disability and/or aged over 65, and engagement – excluding prepayment Excluding prepayment, 45% of the population have annual household income of less than £18,000 and/or a disability and/or are aged over 65			
% of the population who [insert measure of engagement]	32	74	69
% of those who have [insert measure of engagement] with h'hold income of less than £18,000 and/or a disability and/or are aged over 65	57	49	46
% of those with h'hold income of less than £18,000 and/or a disability and/or are aged over 65 who have [insert measure of engagement]	40	81	71
6a. Overlap between those who have annual household income of less than £18,000 and have a disability and aged over 65, and engagement 3% of the population have annual household income of less than £18,000 and a disability and are aged over 65			
% of the population who [insert measure of engagement]	34	75	72
% of those who have [insert measure of engagement] with h'hold income of less than £18,000 and a disability and are aged over 65	5	3	3
% of those with h'hold income of less than £18,000 and a disability and are aged over 65 who have [insert measure of engagement]	52	83	78
6b. Overlap between those who have annual household income of less than £18,000 and have a disability and aged over 65, and engagement Excluding prepayment, 3% of the population have annual household income of less than £18,000 and a disability and are aged over 65			
% of the population who [insert measure of engagement]	32	74	69
% of those who have [insert measure of engagement] with h'hold income of less than £18,000 and a disability and are aged over 65	4	3	3
% of those with h'hold income of less than £18,000 and a disability and are aged over 65 who have [insert measure of engagement]	51	83	75

The SSE Report

Summary of SSE's analysis

6. The SSE analysis uses the data set underlying the customer survey conducted by the company GfK NOP Ltd on the CMA's behalf (the 'GfK Customer Survey').[4] SSE produced a number of tables, estimates of correlations between various variables and a series of logistic regressions to assess the relationship between engagement and customer characteristics. Specifically, it used 21 alternative model specifications in total for its regression analysis, regressing each of the seven dependent variables on the left hand side of Table against each of the three sets of explanatory variables on the right hand side of Table .

[4] This data set combines respondents' responses to the GfK Customer Survey with supplier-provided data. The responses to the GfK Customer Survey covered areas including respondents' history with their supplier, their history of engagement, preferences, activity in other markets, attitudes and demographics. The supplier-provided data included information on customers' tariffs, history with their supplier, consumption and billing information, discounts and schemes or programmes they are part of. See the GFK NOP technical report and Appendix 8.1 of the provisional findings for further detail.

Table 2: Variables used in SSE's regression models

Dependent variables	Explanatory variables	Set 1	Set 2	Set 3
Switched supplier in the last 3 years	Age	Yes	Yes	Yes
Switched supplier in the last year	Household income	Yes	Yes	Yes
Ever switched supplier	Level of qualification	Yes	Yes	Yes
Ever switched supplier or tariff	Whether on PSR	Yes	Yes	Yes
Likely to switch suppliers in the next 3 years	One of carer / disabled / single parent	Yes	Yes	Yes
Confident about being able to make the right decision	Housing tenure	Yes	Yes	Yes
Confident about being able to find the right deal	Internet access	Yes	Yes	Yes
	Whether on WHD	Yes	Yes	Yes
	Contact by other suppliers	Yes	No	Yes
	Contact by own supplier	Yes	No	Yes
	Consider price essential	No	No	Yes
	Consider customer service essential	No	No	Yes
	Consider brand/ reputation essential	No	No	Yes
	Consider other services essential	No	No	Yes

7. SSE's analysis finds that:

(a) demographic factors[5] are correlated with measures of switching;[6]

(b) demographic factors are correlated with each other;

(c) demographic factors are correlated with internet access, receipt of the warm home discount (WHD) and contact by suppliers;

(d) across the specifications internet access; WHD; contact by supplier[7] and housing tenure are either always or generally have a statistically

[5] Age; household income; level of qualifications; whether on the priority services register; being a carer, disabled, or a single parent. We use the term 'demographic factors' as shorthand for these variables throughout the document.
[6] Switched supplier in the last three years; switched supplier in the last year; ever switched supplier; ever switched supplier or tariff.
[7] GfK Customer Survey, question E2A, asked: 'have you ever been approached by your existing energy supplier suggesting you change to a different tariff?' Question E2C asked: 'have you ever been contacted by a different supplier from your current one, suggesting you switched to them?'

significant association with the dependent variable (even after controlling for other demographic factors) whereas other demographic factors are not generally statistically significant.

8. SSE argued that:

(a) It was incorrect to infer that possessing any one of the demographic characteristics would, in itself, make a customer less likely to switch supplier because these characteristics were correlated with:

(i) each other; and

(ii) other factors that might influence switching behaviour. (It identified that these correlated factors may include internet access, WHD and contact by suppliers).

(b) The results of the regression analyses 'indicate that the main drivers of customer switching are internet access, not receiving the WHD and supplier contact' and that demographic factors, except tenure type, had little or no identifiable effect when other factors were controlled for. At its August 2015 hearing, SSE further emphasised that it considered internet access, not receiving the WHD, contact by suppliers and not being in rented accommodation directly influenced, drove, or caused[8]engagement.

(c) The provisional findings had thus failed to identify the main drivers of customer engagement.

CMA comments on SSE's arguments and analysis

9. SSE's arguments set out in paragraph 8 above could be taken to suggest we have identified demographic factors as the main 'drivers' of customer engagement. However, we note that the analysis included in Appendix 8.1 of our provisional findings aimed to generate descriptive statistics and identify material associations between variables in the GfK Customer Survey data set.[9] We did not aim to identify which variables[10] (eg demographics) directly influence other variables (eg switching supplier in the last year). The associations we identified may or may not be due to direct relationships between the variables.

---

[8] We consider these three terms to be synonymous.
[9] Appendix 8.1 covered topics such as customer activity and engagement; gains from switching; reasons for engagement and non-engagement; customers' confidence, capabilities, and experience. It compared results between various subgroups (eg fixed tariffs and SVTs, respondents with incumbent and independent suppliers) and trust.
[10] See footnote 4 for a description of the types of variables in the data set.

10. Identifying associations between variables is valuable because it allows us to identify differences in behaviours, preferences and other relevant variables for subgroups within the population. In Appendix 8.1, we identified many of the associations that SSE found in the data in our own analysis. For example, we remarked on the association between variables related to engagement and both internet access and renting.[11]

11. SSE's analysis showed that there is an association between the likelihood of switching and either the WHD or contact by customers' own supplier or other suppliers. Appendix 8.1 did not report on this association. However, consistent with the SSE results, our further analysis of the GfK Customer Survey data suggests there is a strong association between receiving the WHD and having not switched supplier or tariff.[12] As SSE highlights in its tables, those who receive the WHD are particularly likely to be in demographic groups we identified as being associated with disengagement. This is not surprising given the eligibility criteria for the WHD.[13] The WHD may therefore be expected to be a good predictor of disengagement. Our further analysis of the GfK Customer Survey data is also consistent with there being an association between contact by suppliers and likelihood of switching supplier, although the results are weaker than for the WHD.

12. In addition, the SSE regression analysis provides further evidence that internet access, not receiving the WHD, contact by suppliers, and not renting are positively associated with engagement.

13. Regression analysis can be used to isolate the relationship between a dependent variable and individual explanatory variables from other explanatory variables in the model. The absence of statistical significance between a dependent variable and an explanatory variable (or a set of explanatory variables) may be due to the absence of a direct relationship between them. The existence of a statistically significant association between variables can be consistent with either a direct relationship between variables, an indirect association due to correlation with other factors outside the regression model which influence engagement, or both. Similarly, the magnitude of the measured association can be due to the model capturing direct relationships, indirect relationships or both. If additional assumptions

---

[11] For example, Appendix 8.1, pp 3, 4, 25, 26 & 75.
[12] For example, compared with those who are not in receipt of the WHD, those who do receive it had a lower rate of switching in the last three years (10% compared with 26%), switching in the last year (3% compared with 13%), having ever switched tariff with an existing supplier (19% compared with 29%) and of having ever considered switching (50% compared with 67%).
[13] Customers who receive the Guarantee Credit element of Pension Credit are in the 'core group' and are entitled to receive the WHD if their supplier is part of the scheme. Suppliers who receive means-tested benefits or have a low income may be able to receive the WHD at the discretion of their supplier. See Warm Home Discount Scheme: Eligibility.

are imposed, a model can be said to have mostly or fully isolated and measured the direct relationship between variables. The credibility of such assumptions relies on how reasonable they are given our knowledge of the data and the nature of engagement in the domestic energy market.

14. While there may genuinely be a direct relationship between internet access and renting and customer switching, particularly internet access, it may also be that some of the observed association between these factors and customer switching is due to association with other factors outside the regression model. For example, those who lack internet access may also lack the skills to effectively use the internet to search for and switch suppliers even if they had internet access. SSE has not explained why we should expect its analysis to have isolated the direct relationships. In our view, the observed associations in SSE's results are unlikely to be limited solely to the direct relationship of engagement with internet access and renting. Additionally, it is unclear how much of the observed association can be reasonably ascribed to the direct relationships.

15. SSE argued that the WHD itself directly reduced engagement. In particular, it argued that:

    (a) 'WHD recipients are on favourable tariffs, [and so] one would expect them to switch less than other customers'.

    (b) not all suppliers offered the WHD; and

    (c) supplier discretion in awarding the WHD deterred customers who received it from switching suppliers.

16. However, we note the following:

    (a) The WHD is a rebate rather than a separate tariff.

    (b) The WHD is now offered by the SLEFs, Ovo Energy, First Utility, Co-operative Energy, Utility Warehouse and Utilitia.[14]

    (c) We have not received other evidence that suggests suppliers' discretion over granting the WHD is a substantial source of disengagement among energy customers.

    (d) Based on the GfK Customer Survey data set, there is much higher incidence of receipt of the WHD among [✂] and SSE (dual fuel)

[14] Warm Home Discount Scheme: Eligibility (accessed on 18 November 2015).

customers.[15] As the regression model does not control for supplier, the WHD variable may actually be measuring differences in the likelihood and ability of [✂] and SSE customers to switch relative to customers of other suppliers.

(e) The WHD may be designed to target those least likely to engage in the market.

17. In our view, based on the above, the observed relationship between the WHD and engagement could result primarily from the WHD's association with other factors rather than the WHD itself driving disengagement.

18. Contact by another supplier is likely to directly increase a customer's likelihood of switching or considering switching. Similarly, contact by a customer's own supplier about changing tariff is likely to increase their propensity to do so. However, suppliers also have a commercial incentive to target such communications to those most likely to respond to them and the model will not necessarily disentangle these two effects. It is unlikely that SSE's analysis has isolated the former effect from the latter effect and it is not clear how much of the observed association can be attributed to either.

19. SSE found that individual explanatory variables related to demographic factors did not generally have a statistically significant association with various engagement measures in its regression models. However, we found that this group of variables when considered collectively are statistically significant in each of SSE's 21 specifications. Such a result can occur because the variables are correlated and there is a collective association which the regression analysis is unable to disentangle into individual associations. This result is inconsistent with SSE's finding that demographics have no association with engagement once internet access, WHD, contact by suppliers and renting are controlled for. One possible interpretation of the result is that the analysis does not provide strong evidence of a direct influence of particular demographic factors on engagement but does provide some evidence that the demographic factors are collectively associated with factors outside the model which directly cause variation in engagement.

20. In principle, this type of analysis can be potentially useful for identifying which groups of customers are more likely to be disengaged, whom remedies should be targeted towards and what the remedies should be. The associations SSE observes are either in line with our own findings (internet access, renting) or not surprising given our knowledge of the market (WHD,

[15] These have the lowest rate of switching supplier among the SLEFs for dual fuel customers.

contact by suppliers). In that sense, SSE's results are not particularly useful for our assessment of remedies.

21. Based on the above, our view is that:

(a) SSE's analysis does not invalidate our own analysis of the GfK Customer Survey data; and

(b) the analysis provides evidence that internet access, not receiving the WHD, contact by suppliers and not renting are positively associated with engagement. However, SSE's analysis does not, in our view, reliably isolate and measure the direct influence of these explanatory variables. In particular, the SSE analysis by itself does not provide sufficiently strong evidence to establish that internet access, contact by supplier, not being in receipt of the WHD and not renting are the main factors which directly influence the likelihood of switching and the ability to switch.

The RWE Report

22. The RWE Report contains analyses of gains from switching, unilateral market power over customer groups and product differentiation. We discuss each of these in turn

Gains from switching

Summary of analysis

23. RWE argued that gains from switching (GFS) and observed price variations were at least partly driven by preferences for product characteristics rather than entirely from customer inertia, search costs or switching costs. RWE used an econometric analysis to assess whether GFS differed between those who had and had not engaged when other factors were held constant. It argued that the CMA was not correct to interpret all variation in prices as indicating customer inertia, search or switching costs. RWE suggested that a credible methodology to distinguish between the two competing explanations for the observed price variation would be to examine the gap between the GFS available to those who had and had not recently engaged since that could be interpreted as the GFS due to customer inertia while the remainder of the GFS could be interpreted as the GFS due to customer preferences for tariffs or suppliers with particular characteristics. RWE's analysis relies upon the GfK Customer Survey data set and employs a number of alternative model specifications. The analysis estimates GFS due to inertia on the basis of the estimated value of a coefficient for a particular explanatory variable (or coefficients in some specifications). RWE found that most of the available

GFS were due to preferences rather than inertia. For example, it found that the GFS were only £12 to £23 higher for those who had switched supplier in the last year compared with those who had not.

CMA comments

24. In our view, RWE's analysis has a number of weaknesses. In particular, these include, but are not limited to, the following:

(a) The model uses explanatory variables which capture the characteristics of the current tariff a survey respondent is using and the tariff they would switch to under the relevant GFS scenario. Due to how the GFS estimates are constructed, the GFS available to a respondent will be determined by the tariff they are on, the tariff they would switch to and their consumption level. Therefore, the coefficients for these explanatory variables can be expected to capture most or all of the variation in GFS across survey respondents. This means the coefficients of explanatory variables used to measure GFS due to inertia and relating to the demographic characteristics of survey respondents are not identified and can be expected to be estimated as being close to zero. It is therefore unsurprising that they find the GFS attributable to inertia is small.

(b) The model implicitly assumes that customers who switch, always switch to the tariff which they would most prefer given their preferences. The analysis, as it has been implemented, does not allow for customer errors. In practice, customers may not switch to the optimal tariff given their preferences.[16] The estimates produced by the analysis will therefore tend to overstate the portion of customers' GFS that can be ascribed to preferences.

(c) The model uses a large number of explanatory variables which are correlated. While these factors may control for unobserved heterogeneity, they are also likely to induce multicollinearity problems such as inaccurate and volatile standard errors.[17] This impacts the reliability of findings related to statistical significance.

(d) In some of the specifications used, the GFS attributable to inertia are estimated by comparing the GFS of SVT users who have and have not switched. SVT's status as a default tariff means that many of the observed switches may be due to atypical switching events such as

[16] For example, due to searching over a limited number of tariffs or incorrectly comparing the price or attributes of tariffs.
[17] In this context, multicollinearity is a problem if it affects the variables of interest or variables which are correlated with them.

moving to a new property or due to contact by suppliers. Switching among SVT customers may therefore be a poor basis on which to make inferences about the wider population.

25. Due to these weaknesses, we are unable to place any evidential weight on RWE's analysis.

Analysis of unilateral market power over customer groups

Summary of analysis

26. RWE argued that if suppliers had unilateral market power over specific customer groups, then we should expect to see higher GFS and lower levels of engagement among these groups. RWE used a regression analysis to disentangle the correlations between individual drivers of engagement and propensity to engage. It employed a number of alternative specifications which used different measures of engagement as dependent variables. It used explanatory variables which it argued might be expected to influence engagement including demographics, product characteristics, customer confidence and capability variables, consumption, interactions with tariff types and demographics. It also referred to the results from the GFS analysis with regard to demographics to support its argument.

27. RWE found that whether a customer was on the SVT did not consistently result in a lower likelihood of engagement across all engagement measures once product characteristics, demographics and consumer confidence were controlled for. It has not found the demographic characteristics emphasised by the CMA to all be simultaneously important drivers of engagement and therefore argued that the CMA's provisional findings in this regard were not robust.

CMA comments

28. We are of the view that RWE's analysis has a number of weaknesses. In particular, RWE has used explanatory variables such as tariff type and customer attitudes, which rather than determining engagement, may be actually determined by past engagement.[18] That is, the RWE analysis assumes the wrong flow of causality between the 'dependent' variable and many of the 'explanatory' variables. This will lead to bias in the estimates of these 'endogenous' variables. These 'endogenous' variables are correlated

[18] For example, the association between tariff type and switching may be because people typically switch to fixed tariffs rather than those who are on fixed tariffs being inherently more likely to switch.

with many of the other explanatory variables, and so these other variables will also be contaminated by this bias. This makes it unclear what the estimates are actually measuring and what interpretation should be given to the results.

29. In addition, the following issues pointed out above are also relevant for RWE's analysis:

    (a) The issues highlighted in paragraph 24(c) regarding multicollinearity.

    (b) The issues regarding the GFS results mentioned in paragraph 24(a) regarding the small expected association between demographics and GFS.

    (c) The arguments made regarding the focus of our own analysis in paragraph 9.

30. Due to these weaknesses, we are unable to place any evidential weight on RWE's analysis.

Product differentiation

Summary of analysis

31. In the provisional findings, we said that gas and electricity are homogenous products.[19] RWE has argued that products were differentiated, not homogeneous. Using the tariff data used for the GFS analysis, RWE produced a regression model to estimate how prices varied according to tariff characteristics. It found that tariffs with different characteristics were priced at different levels. Using the GfK Customer Survey data, it provided tables showing that among customers who switched, the incidence of switching to SVTs and online tariffs varied with age.

32. RWE argued that this was evidence that customers being on higher-priced products was due, at least in part, to customer choices which were based on variation in their preferences rather than the only cause being customer inertia. It also argued that the evidence was not consistent with all of the observed price variation reflecting 'disengaged customers suffering from unexploited gains from switching because they are not on the cheapest tariff rate' but '(r)ather the evidence makes clear that customers who have switched also make choices for example to have an online tariff or sometimes to choose an SVT'.

[19] For example, see the provisional findings, paragraph 82.

CMA comments

33. We have not found problems with the underlying analysis and results reported by RWE. The regression analysis is similar but not identical to the analysis conducted by the CMA in Appendix 10.7 of the provisional findings. Its results provide evidence that tariffs with different characteristics tend to have different prices.

34. RWE's analysis by itself cannot identify whether price differentials are due to customer preferences, cost differences across tariffs and suppliers, variation in pricing and customer segmentation strategies by suppliers or some combination of these factors. Additionally, it is not obvious why an association between age and the types of tariffs customers switch to should necessarily undermine a finding that disengaged customers could gain from switching.