Appendix H: Prevalence of switching methodology

H.1 This appendix sets out the data, methodology and results of our analysis on the prevalence of switching. Specifically, we consider the prevalence of full switching in 2020, 2021 and 2022 between customers of AWS, Microsoft and Google.

Data

H.2 AWS, Microsoft and Google each provided us with customer data sets with customer names, years and annual revenues.[1] The data requested specified customers with an annual spend over $1,000 in each of the given years.[2]

H.3 AWS provided customer data for 2018-2023. Microsoft's data set covers 2018-2023. Google only provided data for 2020-2023. This results in the shared years in the data set for all three providers being 2020-2023.

H.4 We do not have similar quality data from any other cloud providers. This means if customers were switching to or away from other providers, eg Oracle, IBM, etc, we are not able to capture them. However, Oracle and IBM make up a relatively small share of the market (see Chapter 2 market shares and concentration), so we expect this limitation to have a limited impact on our results.

Methodology

H.5 In order to identify customers who have fully switched from one cloud provider to another we took the following steps:

(a) First, we identified customers who have potentially fully switched away from one cloud provider.

(b) Second, we matched customer data sets across providers to identify any such customers that are present in multiple providers' customer data sets;

(c) Third, we checked if those customers subsequently increased their spend with another cloud provider. If so, we identified them as full switchers.

[1] Responses to the CMA's information requests [✂].
[2] [✂].

H.6 Below, we elaborate on each of these steps and describe some limitations which affect the interpretation of our results.

Identifying potential switchers

H.7 Our first step was to identify customers who have left a cloud provider. These are customers who reduced their spend with one cloud provider significantly. We allowed for this spend not to fall completely to zero to capture cases where some residual marginal service is left in place with the original provider.

H.8 In particular, we identified potential switchers as those customers whose spend with one provider has fallen by 85% year-on-year. We performed sensitivity tests around this threshold.

Matching customers across providers

H.9 The next step was to match the identified potential switchers to customers in other providers' data sets.

H.10 In order to identify customers common across different providers' data sets, we adopted a similar methodology to the one we used for our analysis of the prevalence of multi-cloud (see Appendix I).

H.11 We matched customers' names across the customer data sets from AWS, Microsoft and Google. We used two types of matching:

(a) Perfect matching: exact matches of customer names across data sets.

(b) Fuzzy matching: matches based on similar but non-identical strings in customer names. Fuzzy matching produces a similarity score based on how good the match is, with 0 meaning the two are not a match and 1 meaning a perfect match. We chose to use fuzzy matching to capture additional matches where customer names may have been recorded slightly differently across the providers' data sets (eg 'Company A' in one data set but 'Company A LTD' in another).

H.12 For fuzzy matches, we set a cut-off for the similarity score of 99%: under this level of similarity, we do not consider fuzzy matches to be reliable enough based on cross-checks that we performed. We selected this threshold after manually checking a sample of matches with different levels of similarity scores and looking for false positives (ie different customers being matched) and false negatives (ie the same customer not being matched). See Appendix I for a more detailed explanation of this process.

Identifying actual switchers

H.13 Once the potential switchers had been matched across data sets, we determined whether they are actually switchers or not. To do that, we set the following condition: a potential switcher is identified as an actual switcher only if it increases its spend on another provider by at least 60% of the amount it reduced its spend on the original provider.

H.14 For example, if Customer A decreased spend on Azure by $10,000 and increased their spend on AWS by $6,000 or more, they would be identified as an actual switcher.[3]

H.15 We performed sensitivity tests around this threshold.

Caveats and limitations

H.16 In this section, we describe the main limitations and caveats to our analysis. We also present how these limitations and caveats affect the interpretation of the results.

Coverage of providers

H.17 Our data covers only the three largest providers. This means we could not capture switching from or to smaller providers like Oracle, IBM and others. While the three largest providers cover most of the market (see Chapter 3), we acknowledged that this may lead to underestimating the switching in the market.

Time Frame

H.18 Given the limitations of the data sets, specifically only having three years where we could observe switching, we have decided that one year is a suitable timeframe for switching.[4] That is, we assessed changes in spending on a year-by-year basis.

H.19 We acknowledge that switching, especially for large customers, can be a very complicated and time intensive project. To the extent that switching occurs over multiple years, our methodology might not capture such occurrences.

---

[3] A customer who spends 60% of their decrease across multiple cloud providers are also identified as an actual switcher. If Customer A decreased spend on Azure by $10,000 and increased spend on AWS and Google by $3,000 each (ie 30% + 30%); they would be identified as an actual switcher.

[4] We have four years of spend data. However, to observe switching we need to see a decrease in spend in year 1 coupled with an increase in spend in year 2. This means we cannot identify switching from the last year in the data.

Switching criteria

H.20 As described above, we made some assumptions on the year-on-year changes in spend in order to identify switchers. We performed sensitivities around these assumptions to make sure they do not drive our results. We present these sensitivities alongside our results.

Results

H.21 In this section, we present the results of our analysis.

H.22 Table H.1 shows the total revenue that switched away from the top three providers from 2020 to 2022.[5] The table suggests that the percentage of revenue that switched between providers is very limited, below 0.3% over the three years. Note that our definition of 'switched away' reflects the methodology described above, ie only matched customers who fulfil all our criteria are described as switchers.

Switched customers and revenue

Table H.1: Sum of revenues switched 2020-2022

Year	*Switched revenue*	*Total revenue*	*Percentage*
2020	$ 1,841,984	$ 3,748,034,240	0.049%
2021	$ 2,876,135	$ 5,224,041,088	0.05%
2022	$ 18,478,412	$ 7,146,011,648	0.259%

Source: CMA analysis of parties' data

H.23 Over 50% of the switched revenue in 2022 comes from a single switching customer. Without that customer the percentage of revenue that switched away in 2022 would be roughly 0.1%.

H.24 Table H.2 shows the number of customers (unweighted) who switched in each year and the total number of customers across our data sets in each year. In each year only about 0.9% of all customers switched.

Table H.2: Customers that switched cloud provider, 2020-2022

Year	Switching customers	Total customers	Percentage
2020	342	37,279	0.92%
2021	370	43,720	0.84%
2022	430	50,019	0.86%

Source: CMA analysis of parties' data

[5] We have four years of spend data. However, to observe switching we need to see a decrease in spend in year 1 coupled with an increase in spend in year 2. This means we cannot identify switching from the last year in the data

H.25 Table H.1 and Table H.2 together indicate that there are very low levels of switching amongst the top three providers. That being said, both tables are likely an underestimation of both the total amount of revenue switched and the unweighted number of customers that switched, as not all accurate fuzzy matches will have been captured at our threshold of 99%. Similarly, there would likely have been customers who switched to or from providers not in our data sets that would have increased our estimates.

Distribution of switchers

H.26 We also looked at the size of those customers we identified as switchers. In particular, we note that the unweighted number of switchers is higher than their weighted counterpart, which suggests that most switchers have smaller than average cloud expenditures.

H.27 Figure H.1 below shows the distribution of switchers by their spend size (ie the amount of spend they switched).

Figure H.1: Distribution of switchers by size of revenue switched

[✂]

H.28 Figure H.1 suggests that approximately 80% of all switchers spend between $[✂] - $[✂] on cloud providers. These customers may be more likely to switch because:

(a) They have simpler cloud systems.

(b) They may be trialling the cloud providers to assess suitability for their needs.

(c) Less likely to encounter commercial barriers to switching.

H.29 There is a notable lack of switching amongst larger spending customers. The largest customer moved $[✂] from one provider to another. Including this customer, only three customers switched spends greater than $[✂].

H.30 These larger customers are more likely to have complex systems that require time and considerable funds to move between providers.[6] The complexity of these systems might mean that a one year switching window may be insufficient and narrow.

[6] [✂] response to the CMA's information requests [✂].

Sensitivities

H.31 In setting our threshold to identify switchers we examined the sensitivity of the results to different thresholds. We examined the number of customers and the revenue identified as switching at every combination of thresholds; from -50% decrease and 50% increase, to -100% decrease with a 100% increase.

H.32 In doing this we found a steep drop-off in revenue after the -85% decrease threshold. We therefore decided to set the decrease threshold at -85%. We also found a steep drop off after the 80% increase threshold. This was less relevant as we felt a lower increase threshold was more appropriate. We found no steep drop off in the number of customers at any threshold level.

Our views on one provider's switching analysis

H.33 One provider submitted estimates of customer switching. We consider that there are limitations with this analysis[7] which we set out below:

(a) The provider defined customers as churned if the customer reduced its spending by at least 80% for three consecutive months and did not return to 80% of their original spend within six months. This may only imperfectly capture the actual churn – a customer may still continue to buy from that provider and/or significantly increase its spending on that provider's services after six months. The provider's measure of churn may reflect normal seasonal fluctuation – it is difficult to draw conclusions from the short time frame.

(b) The data set includes many small customers who only spend a minimal amount on the provider's cloud services during the entire period considered. These customers may only have been trialling the provider's services or only used the services occasionally. We note that:

(i) Removing customers that spent less than $100 reduces the churn rate from [✂]% to [✂]%.

(ii) Removing customers that spent less than $500 reduces the churn rate to [✂]%.

(c) Customer spend in the data set is heavily skewed towards the largest customers. The top 5% of the provider's customers account for [✂]% of the provider's total revenue and the top 10% account for [✂]%. The churn rate varies for different customer spend deciles:

[7] [✂] response to the CMA's information requests [✂].

(i) For customers above the 7th decile in spend, the churn rate is below [✂]% in each decile.

(ii) For customers in the top 10% of spend, the churn rates are approximately [✂]%.

H.34 With regard to the provider’s calculations on customers that decreased spending between the first half of 2020 and the first half of 2022, we consider that a reduction in spending alone is not a meaningful measure of switching. There are many reasons a customer may have reduced spending; it does not necessarily follow that the customer is switching away all or part of its workloads.

H.35 For these reasons, we consider that the results from the provider’s quantitative analysis should be interpreted in light of these caveats. We note that, in any case, switching rates calculated by the provider of [✂]% and [✂]% are not inconsistent with finding low switching levels in the cloud infrastructure services market.