Appendix: proving the CACE formula

Proving the CACE formula is a bit cumbersome and convoluted, but there is no actual difficulty. Let's use the following abbreviations for readability:

- Always-takers = AT
- Never-takers = NT
- Compliers = CMP
- Treatment group = TG
- Control Group = CG

I'll also use the conditional operator "|" to indicate "in", i.e. "|TG" means "in the treatment group".

We want to calculate the difference between the average booking profit per day (noted *aBP* for short) of compliers in the treatment group and the *aBP* of compliers in the control group: $CACE = aBP(CMP|TG) - aBP(CMP|CG)$.

We can express the *aBP* in the control group as the *aBP* for treated (a.k.a. always-takers), weighted by their proportion in the control group, plus the *aBP* for untreated (a.k.a. compliers and never-takers, in unknown proportions at this point), similarly weighted

$$aBP(CG) = aBP(treated|CG) * P(treated|CG) + aBP(untreated|CG) * P(untreated|CG)$$

Similarly, the *aBP* in the treatment group (TG) is the *aBP* for untreated (a.k.a. never-takers) plus the *aBP* for treated (a.k.a. compliers and always-takers in unknown proportions), multiplied by their respective weights

$$aBP(TG) = aBP(untreated|TG) * P(untreated|TG) + aBP(treated|TG) * P(treated|TG)$$

1. Calculating $aBP(CMP|TG)$

While we don't know (yet) the corresponding numerical values, we can conceptually express the *aBP* for compliers and always-takers in the treatment group, using hats for unknown values:

$$aBP(treated|TG) = aBP(\widehat{CMP}|TG) * \frac{P(\widehat{CMP|TG})}{P(treated|TG)} + aBP(\widehat{AT}|TG) * \frac{P(\widehat{AT|TG})}{P(treated|TG)}$$

Isolating the variable we're looking gives us:

$$aBP(\widehat{CMP}|TG) = \left[aBP(treated|TG) - aBP(\widehat{AT}|TG) * \frac{P(\widehat{AT|TG})}{P(treated|TG)}\right] / \frac{P(\widehat{CMP|TG})}{P(treated|TG)}$$

Let's assume that per randomization the proportion and *aBP* of always-takers in the treatment group are the same as in the control group (which we know):

$$P(\widehat{AT|TG}) = P(AT|CG) = P(treated|CG)$$

$$aB\widehat{P(AT|}TG) = aBP(AT|CG) = aBP(treated|CG)$$

Then we can deduce that:

$P(\widehat{CMP|TG}) = P(treated|TG) - P(treated|CG)$

Replacing all the unknowns by their expressions in the previous equation yields:

$$aBP(\widehat{CMP}|TG) = \left[aBP(treated|TG) - aBP(treated|CG) * \frac{P(treated|CG)}{P(treated|TG)}\right] / \frac{P(treated|TG)-P(treated|CG)}{P(treated|TG)}$$

Let's simplify:

$$aBP(\widehat{CMP}|TG) = aBP(treated|TG) * \frac{P(treated|TG)}{P(treated|TG)-P(treated|CG)} - aBP(treated|CG) * \frac{P(treated|CG)}{P(treated|TG)-P(treated|CG)}$$

2. Calculating $aBP(CMP|CG)$

Let's apply the exact same logic in the control group (it's literally copy-pasted, just changing the variable names): while we don't know (yet) the corresponding numerical values, we can conceptually express the *aBP* for compliers and never-takers in the treatment group, using hats for unknown values:

$$aBP(untreated|CG) = aBP(\widehat{CMP}|CG) * \frac{P(\widehat{CMP}|CG)}{P(untreated|CG)} + aBP(\widehat{NT}|CG) * \frac{P(\widehat{NT}|CG)}{P(untreated|CG)}$$

Isolating the variable we're looking gives us:

$$aBP(\widehat{CMP}|CG) = \left[aBP(untreated|CG) - aBP(\widehat{NT}|CG) * \frac{P(\widehat{NT}|CG)}{P(untreated|CG)}\right] / \frac{P(\widehat{CMP}|CG)}{P(untreated|CG)}$$

Let's assume that per randomization the proportion and *aBP* of never-takers in the control group are the same as in the treatment group (which we know):

$P(\widehat{NT|CG}) = P(NT|TG) = P(untreated|TG)$

$aBP(\widehat{NT|CG}) = aBP(NT|TG) = aBP(untreated|TG)$

Then we can deduce that:

$P(\widehat{CMP|CG}) = P(untreated|CG) - P(untreated|TG)$

Replacing all the unknowns by their expressions in the previous equation yields:

$$aBP(\widehat{CMP}|CG) = \left[aBP(untreated|CG) - aBP(untreated|TG) * \frac{P(untreated|TG)}{P(untreated|CG)}\right] / \frac{P(untreated|CG)-P(untreated|TG)}{P(untreated|CG)}$$

Let's simplify:

$$aBP(\widehat{CMP}|CG) = aBP(untreated|CG) * \frac{P(untreated|CG)}{P(untreated|CG)-P(untreated|TG)} - aBP(untreated|TG) * \frac{P(untreated|TG)}{P(untreated|CG)-P(untreated|TG)}$$

Let's rewrite all the proportions of untreated as one minus the proportion of treated, for consistency with the formula for the treatment group:

$$aBP(\widehat{CMP}|CG) = aBP(untreated|CG) * \frac{1-P(treated|CG)}{(1-P(treated|CG))-(1-P(treated|TG))} - aBP(untreated|TG) * \frac{1-P(treated|TG)}{(1-P(treated|CG))-(1-P(treated|TG))}$$

We can simplify out the ones in the denominators of the fractions:

$$aBP(\widehat{CMP}|CG) = aBP(untreated|CG) * \frac{1-P(treated|CG)}{P(treated|TG)-P(treated|CG))}$$
$$- aBP(untreated|TG) * \frac{1-P(treated|TG)}{P(treated|TG)-P(treated|CG)}$$

3. Simplifying the difference

$$CACE = aBP(CMP|TG) - aBP(CMP|CG) =$$
$$\left[aBP(treated|TG) * \frac{P(treated|TG)}{P(treated|TG)-P(treated|CG)} - aBP(treated|CG) * \frac{P(treated|CG)}{P(treated|TG)-P(treated|CG)}\right] -$$
$$\left[aBP(untreated|CG) * \frac{1-P(treated|CG)}{P(treated|TG)-P(treated|CG))} - aBP(untreated|TG) * \frac{1-P(treated|TG)}{P(treated|TG)-P(treated|CG)}\right]$$

Let’s bring out the denominator:

$$CACE = \frac{1}{P(treated|TG)-P(treated|CG)} *$$
$$\{[aBP(treated|TG) * P(treated|TG) - aBP(treated|CG) * P(treated|CG)] -$$
$$[aBP(untreated|CG) * (1 - P(treated|CG)) - aBP(untreated|TG) * (1 - P(treated|TG))]\}$$

Simplify out the straight brackets:

$$CACE = \frac{1}{P(treated|TG)-P(treated|CG)} *$$
$$\{aBP(treated|TG) * P(treated|TG) - aBP(treated|CG) * P(treated|CG) -$$
$$aBP(untreated|CG) * (1 - P(treated|CG)) + aBP(untreated|TG) * (1 - P(treated|TG))\}$$

And

$$CACE = \frac{1}{P(treated|TG)-P(treated|CG)} *$$
$$\{[aBP(treated|TG) * P(treated|TG) + aBP(untreated|TG) * (1 - P(treated|TG))] -$$
$$[aBP(treated|CG) * P(treated|CG) + aBP(untreated|CG) * (1 - P(treated|CG))]\}$$

But the term between the curly brackets is simply the ITT:

$$CACE = \frac{ITT}{P(treated|TG)-P(treated|CG)}$$

Which is the formula in the body of the chapter.