A proof that using weighted least squares, both Karlson's weight and Zhou's weight will produce unbiased estimates.

Before we get to it, first note that it can be proved that Zhou's weight can be directly used in the weighted Y approach, if both X and C are binary. Concretely, $E(Y(C=1)|X=1) - E(Y(C=1)|X=0) - [E(Y(C=0)|X=1) - E(Y(C=0)|X=0)]$ can be unbiasedly estimated simply by the difference-in-mean estimator, i.e. the sample analogy of $E(YW_{Zhou}|C=1, X=1) - E(YW_{Zhou}|C=1, X=0) - E(YW_{Zhou}|C=0, X=1) + E(YW_{Zhou}|C=0, X=0)$. The proof below, is for using WLS, instead of this simple difference-in-mean estimator, for estimation, whereby both weights produce unbiased estimates.

To clarify, the causal estimand of interest, in general, is $E(Y_C|X)$. In words, this is the association between observed X and potential outcomes Y_C.

The proof here borrows heavily from Robins (1999) in Synthese. Similar steps also spear in Robins (2000) chapter in Statistical Models in Epidemiology, the Environment, and Clinical Trials.

Define $f(Y_C, C, X, Z)$ to be the joint density that produces the factual and counterfactual data. We can factorize it:

$$\begin{aligned}&f(Y_C, C, X, Z)\\&= f(Y_C|X)f(C|Y_C, Z, X)f(Z|Y_C, X)f(X)\\&= f(Y_C|X)f(C|Z, X)f(Z|Y_C, X)f(X)\end{aligned}$$

The second equality is by the identifying assumption $Y_C \perp C|Z, X$.

Now I first prove the unbiasedness of Zhou's weight. We define a new joint density $f^*(Y_C, C, X, Z)$. And this new density function, by definition, has $f^*(C|Y_C, Z, X) = f^*(C|Z, X) = f^*(C|X) = f(C|X)$. Also, by definition, this new density function only differs from the original density in terms of $f^*(C|Y_C, Z, X)$.

Based on how we define the new density function, $f^*(Y_C, C, X, Z)$ can be written as

$$\begin{aligned}&f^*(Y_C, C, X, Z)\\&= f^*(Y_C|X)f^*(C|Y_C, Z, X)f^*(Z|Y_C, X)f^*(X)\\&= f^*(Y_C|X)f(C|X)f^*(Z|Y_C, X)f^*(X)\\&= f(Y_C|X)f(C|X)f(Z|Y_C, X)f(X)\end{aligned}$$

Note that first, our estimand is produced by $f(Y_C|X)$, which is unchanged in the new joint density function. This means data, factual or counterfactual, produced from $f^*(Y_C, C, X, Z)$ will have the same values for the estimand of our interest as in the original data. Second, under $f^*(Y_C, C, X, Z)$, $Y_C \perp C|X$. A brief proof of the second point is as follows: $Y_C \perp^* C|X \asymp C \perp^* Z|X$. This is the equivalence of causal exogeneity and statistical exogeneity, the latter means C is not correlated with Z conditional on X. Note that unobserved confounders are assumed away. And $C \perp^* Z|X$ directly follows from the definition $f^*(C|Z, X) = f^*(C|X)$.

Therefore, $E(Y_C|X) = E^*(Y_C|X) = E^*(Y_C|X, C) = E^*(Y|X, C)$. As the standard OLS mechanics indicate, solving for $q(X, C)$ in the normal equation $E^*\{q(X, C)[Y - q(X, C)]\} = 0$ will give an unbiased estimate of the CEF $E^*(Y|X, C)$, when $q(X, C)$ is correctly specified in X and C. The correct specification is important, as only when the functional form in the regression model agrees with the functional form in the true CEF will the condition required for unbiasedness, $E^*(Y - q(X, C)|X, C) = 0$, kick in. See page 48 and 32 in mostly harmless.

Further,

$$
\begin{aligned}
&E^{*}\{q(X,C)[Y-q(X,C)]\} \\
&= \iiint q(X,C)[Y-q(X,C)]f^{*}(Y,X,C)dydxdc \\
&= \iiint\int q(X,C)[Y-q(X,C)]f^{*}(Y_C,X,C,Z)dydxdcdz \\
&= \iiint\int q(X,C)[Y-q(X,C)]f^{*}(Y_C,X,C,Z)dydxdcdz \\
&= \iiint\int q(X,C)[Y-q(X,C)]\frac{f^{*}(Y_C,X,C,Z)}{f(Y_C,X,C,Z)}f(Y_C,X,C,Z)dydxdcdz \\
&= E\left\{q(X,C)[Y-q(X,C)]\frac{f^{*}(Y_C,X,C,Z)}{f(Y_C,X,C,Z)}\right\} \\
&= E\left\{q(X,C)[Y-q(X,C)]\frac{f(C|X)}{f(C|X,Z)}\right\}
\end{aligned}
$$

Hence, solving for $q(X,C)$ in $E\left\{q(X,C)[Y-q(X,C)]\frac{f(C|X)}{f(C|X,Z)}\right\}=0$ is equivalent to solving for $q(X,C)$ in $E^{*}\{q(X,C)[Y-q(X,C)]\}=0$. And the regression estimand for the estimated $q(X,C)$ is $E^{*}(Y|X,C)$, which we have shown to be equal to $E(Y_C|X)$, our causal estimand. Also, the sample analog of $E\left\{q(X,C)[Y-q(X,C)]\frac{f(C|X)}{f(C|X,Z)}\right\}$ is just the estimating equation for the WLS with weight $\frac{P(C|X)}{P(C|X,Z)}$. The proof is thus completed.

Now I move on to Karlson's weight. Define yet another new joint density function $f^{**}(Y_C,C,X,Z)$, and $f^{**}(Y_C,C,X,Z)=f^{**}(Y_C,C,X)$. This equality implies $f^{**}(Z|Y_C,C,X)=f^{**}(Z)=1$, which means Z has a degenerate density hence is independent of Y_C,C,X. In concrete terms, in this newly defined joint density, Z has only one value with non-zero probability. More simply put, Z is a constant now. Robins also invokes the concept of degenerate density in his 2000 chapter in a discussion on the unstablized IPTW. We also define $f^{**}(Y_C,C,X,Z)$ in a way that makes $f^{**}(Y_C|X)=f(Y_C|X)$, $f^{**}(C|Y_C,X)=f(C|Y_C,X)$, and $f^{**}(X)=f(X)$.

Similar to before, we factorize the new joint density:

$$
\begin{aligned}
&f^{**}(Y_C,C,X,Z) \\
&= f^{**}(Y_C,C,X) \\
&= f^{**}(Y_C|X)f^{**}(C|Y_C,X)f^{**}(X) \\
&= f(Y_C|X)f(C|Y_C,X)f(X)
\end{aligned}
$$

Same as the previous case, the estimand-generating conditional density $f^{**}(Y_C|X)$ is unchanged compared with the original conditional density. And $Y_C \perp^{**} C|X$ is guaranteed by $C \perp^{**} Z|X$, which directly follows from the fact that Z is defined to be a constant under $f^{**}(Y_C,C,X,Z)$. Hence, $E(Y_C|X)=E^{**}(Y_C|X)=E^{**}(Y_C|X,C)=E^{**}(Y|X,C)$.

Again, we estimate $E^{**}(Y|X,C)$ with the estimating equation $E^{**}\{q(X,C)[Y-q(X,C)]\}$. As before,

$$
E^{**}\{q(X,C)[Y-q(X,C)]\}
$$

$$= E\left\{q(X,C)[Y-q(X,C)]\frac{f^{**}(Y_C,X,C,Z)}{f(Y_C,X,C,Z)}\right\}$$
$$= E\left\{q(X,C)[Y-q(X,C)]\frac{1}{f(C|X,Z)}\right\}$$

Therefore, estimating a WLS with weight $\frac{1}{P(C|X,Z)}$ will also lead to an unbiased estimate of $E(Y_C|X)$.

Comment: 1) now it becomes clear that weights serving as unbiased estimator of $E(Y_C|X)$ are not unique. In fact, the numerator can be any function of C. 2) as we know in the classic case of OLS, correctly specifying the regression function $q(X,C)$ is important. If the underlying CEF $E^{*}(Y|X,C)$ or $E^{**}(Y|X,C)$ is not linear, the estimated linear $q(X,C)$ will only be a best linear approximation of the true CEF.