accesses to clean water. The evidence of propensity score misspecification is weaker when using the KS_n test statistic (2.9). *Spec2*, on the other hand, is not rejected for any outcome at the usual significance levels, using either CvM_n or KS_n test statistic. Thus, our tests suggests that *Spec2* should be preferred when analyzing per capita CO_2, deforestation and energy depletion, whereas for urban and rural water access, our tests do not favor either specification.

Table 4: Effect of GATT/WTO membership on environmental quality

	Per capita CO_2		Deforestation		Energy Depletion		Rural water access		Urban water access	
	*Spec*1	*Spec*2	*Spec*1	*Spec*2	Spec 1	*Spec*2	*Spec*1	*Spec*2	*Spec*1	*Spec*2
$\widehat{ATE_n}$	-1.29	-1.00	0.26	0.34	-3.35	-3.39	3.07	2.89	-5.21	-4.62
	(0.58)	(0.49)	(0.20)	(0.21)	(1.35)	(1.38)	(4.84)	(4.79)	(3.75)	(3.72)
	[0.025]	[0.039]	[0.203]	[0.105]	[0.013]	[0.014]	[0.526]	[0.547]	[0.165]	[0.213]
$\widehat{ATT_n}$	-0.81	-0.56	0.14	0.22	-2.16	-1.62	1.74	1.33	1.74	-3.34
	(0.68)	(0.59)	(0.20)	(0.21)	(1.30)	(1.34)	(5.02)	(4.98)	(4.19)	(4.16)
	[0.234]	[0.338]	[0.500]	[0.287]	[0.097]	[0.228]	[0.730]	[0.790]	[0.679]	[0.422]
CvM_n	0.01	0.28	0.01	0.45	0.01	0.35	0.53	0.26	0.49	0.23
KS_n	0.12	0.24	0.17	0.57	0.12	0.43	0.70	0.52	0.63	0.28

Note: *Spec*1 and *Spec*2 are different specifications of the propensity score. $\widehat{ATE_n}$ and $\widehat{ATT_n}$ are the estimators for ATE and ATT in (4.3) and (5.1), respectively, but with observations with estimated propensity score outside $[0.05, 0.95]$ trimmed. Standard errors are in parenthesis, and p-values in brackets. "CvM_n" and "KS_n" respectively stand for the bootstrapped p-values of our proposed Cramér-von Mises and Kolmogorov-Smirnov tests based on 100,000 bootstrap draws. See the main text for further details.

Next we comment on the consequences of propensity score misspecification. For per capita CO_2, our results suggest that the overall effect of GATT/WTO membership on emissions is negative and statistically significant at the 5% level under both propensity score specifications. On the other hand, we find the effect of GATT/WTO membership on per capita CO_2 is not statistically significant among the treated sub-population using either specification. In terms of point estimates, however, there are important differences. For example, the ATE point estimate under *Spec1* (misspecified propensity score) is 30% higher (in absolute terms) than under *Spec2*. Note that the 0.3 difference in ATE represents roughly 8% of the overall per capita CO_2 emissions.

When we analyze the effect of GATT/WTO on deforestation and energy depletion, our results again highlight the consequences of propensity score misspecifications. We find that the ATE point estimate for the effect of GATT/WTO membership on deforestation is 30% larger under *Spec2* than under *Spec1*, and the ATT point estimate for the effect of GATT/WTO membership on energy depletion is 25% smaller under *Spec2* than under *Spec1*. Such large differences are economically significant, as the 0.08 difference in $ATEs$ on deforestation represents nearly 12% of the mean annual deforestation, and the 0.46 difference in $ATTs$ on energy