POLITICAL ECONOMY RESEARCH INSTITUTE

PERI

University of Massachusetts Amherst

POLITICAL ECONOMY RESEARCH INSTITUTE

Not So Clear:
Revisiting the Impacts of Cap-and-Trade on Environmental Justice

Michael Ash and Manuel Pastor

June 2023
Updated July 2023

WORKINGPAPER SERIES

Number 579

Not So Clear:

Revisiting the Impacts of Cap-and-Trade on Environmental Justice

By Michael Ash and Manuel Pastor

Draft: July 20, 2023

Abstract: A recent article by Danae Hernandez-Cortes and Kyle Meng (HCM) suggests that the cap-and-trade program in California was accompanied by improvements in the degree of environmental inequity in the state. We note that that the model used to estimate this improvement is not well-designed to capture the variation in facility adjustment to the cap-and-trade program that is at the heart of the environmental justice debate about potential shifts in co-pollutant exposure. We also show that even if that were a proper approach, the estimates offered by HCM may be problematic due to data issues, including proper identification of facilities subject to the cap, shifting results when we require that facilities have observations both before and after the cap, and robustness when we apply their method estimates beyond their selected subsample to the broader range of facilities. As such, the environmental justice implications of California's carbon market remain an unsettled empirical question.

Introduction

The fundamental selling-point of market-based environmental instruments, both taxes and permits, is that they allow polluters to choose different rates of abatement based on different marginal costs of abatement, with the promise of achieving a given systemwide total abatement at least cost (Stavins 1998). However, this quality of enabling discretionary differences, in contrast to environmental regulations that mandate identical across-the-board reductions in pollution, is a concern for environmental justice (EJ) scholars and advocates who worry about potential variations in co-pollutant reduction (Boyce and Pastor 2013).

While the issue of cap-and-trade is frequently the subject of ideological debates, a new article "Do Environmental Markets Cause Environmental Injustice? Evidence from California's Carbon Market" by Hernandez-Cortez and Meng (2023, henceforth HCM) rightly notes that the question of whether carbon markets improve or worsen current environmental disparities in exposure is hard to predict from theoretical considerations alone and hence it is the facts that matter. In their article, HCM highlight their introduction of a comparison between cap-and-trade and non-covered facilities and the application of air modeling as improvements over previous approaches, including the geographic proximity analyses that we and others have utilized in looking at relative air pollution burdens from market-based strategies (Cushing et al. 2018; Pastor et al. 2022; Plummer et al. 2022). They conclude that carbon markets in California led to a reduction of an environmental justice (EJ) gap.

We raise here two concerns about HCM's analysis. The first is whether the method HCM employ to estimate the effect of cap-and-trade can answer the question of variation among regulated facilities at the heart of the policy debate. The second has to do with a range of data issues, including the classification of regulated and non-regulated facilities, the unbalanced panel and identification of trends, and the restricted sample of facilities, all of which call into question the internal and external validity of the HCM findings. The analytical and data issues in HCM leave open the question of whether California's environmental market experiment was associated with improvements or worsening of environmental injustice.

Do HCM Ask the Right Question?

One of the issues raised by environmental justice (EJ) critics of cap-and-trade (C&T), or indeed both tax and permit systems, is that allowing firms to decide whether to curtail emissions or, effectively, to pay others to reduce, will necessarily lead to uneven geographic reductions. Such unevenness by space matters little with regard to greenhouse gas emissions (GHGs) – a reduction anywhere is of general benefit – but it could matter with regard to local co-pollutants, such as PM, NOX, or other hazardous air pollutants that are often correlated with the GHG emissions in both levels and changes (Boyce and Pastor 2013; Zwickl, Sturn, and Boyce 2021) .

A geographically differentiated pattern of emissions reduction is actually the intent of a carbon market: one wants to wring efficiency out of imposing GHG standards by allowing some firms who can more easily and cheaply meet reduction goals to do so while others facing higher barriers to change seek to forego (or minimize) such reductions. The environmental justice gap thus depends on the extent of the total reduction *and* the pattern of heterogeneity in changes, particularly with regard to impact on EJ communities.

HCM's approach starts by attempting to identify the effect of the California carbon permit system, estimating a common percentage reduction across regulated facilities relative to non-regulated facilities. To get at heterogeneity, in a second stage, HCM then converts these assumed constant-percentage effects into predicted changes in physical quantities of pollutants and applies an air model to estimate an effect on the aggregate pollution disparities between the bulk of California neighborhoods and Disadvantaged Communities (DACs). DACs are the California regulatory term for communities assessed by the state's CalEnviroScreen tool to be socially vulnerable and overexposed to environmental hazards.

To estimate the average effect, HCM utilize the following equation on an unbalanced panel of covered and non-covered facilities observed before and after the policy implementation to obtain predicted values that they then feed into their air model:

Equation 1: $$asinh\ Y_{jt}^{p} = \kappa_1^{p}[C_j \times t] + \kappa_2^{p}[C_j \times 1(t \geq 2013) \times t] + \Phi_j^{p} + \gamma_t^{p} + \mu_{jt}^{p}$$

where p refers to the pollutant, j refers to facility, C_j is an indicator for whether facility j is covered by the C&T regulation, t is the year, Φ_j^{p} is the facility fixed effect, γ_t^{p} is a year fixed effect (to control for changing macroeconomic, demand, and technological conditions), κ_1^{p} is the differential 2008-2012 pre-trend for the cap-and-trade facilities before the policy actually went into effect, κ_2^{p} is the key estimate of the differential trend after 2012 (2013-2017) when cap-and-trade regulations went into effect, and μ_{jt}^{p} is the error term.[1]

While a trend break,, κ_2^{p}, that is negative would signal some relative improvement to the pre-policy trend, key to the finding of an *absolute* reduction in estimated pollution under C&T is the differential post-trend estimate ($\kappa_1^{p} + \kappa_2^{p}$); in Table 1 of HCM (2023), all these estimates for various pollutants are reported to be negative. Note that what is required for this to be true is that the trend break, κ_2^{p}, be negative of a large enough magnitude to overwhelm the pre-trend which is uniformly estimated by HCM to be positive (that is, relative pollution from the C&T facilities is estimated to have been rising prior to C&T).[2]

Note, moreover, that what is being estimated (after controlling for year and facility fixed effects) is a common percentage effect – in short, the world is modeled as having the highly unlikely feature that every regulated firm adjusts to the imposition of C&T by exactly the same percent. The logic seems to be that this allows for the researcher to deduce a "pure" C&T effect since the actual data might include noise due to other factors. But in this case, divergence in outcomes does not emerge only from random dispersion around a mean but occurs because the intent of the policy design is differential adjustment. An estimating strategy that rules that out by design seems misplaced.

This search for a precise C&T policy effect also leads the authors to reduce their sample to a sharply restricted set of facilities. They specifically eliminate refineries and electricity generators because they were subject to other confounding regulations; because selection into the treated group depends on crossing a threshold for GHG emissions, they also drop other large facilities to "ensure better comparability between treated and control facilities" (Hernandez-Cortes and Meng 2023:6). HCM winds up considering about five percent of the GHGs regulated by the program. While the authors report that they later test whether an analysis of the whole dataset finds similar results in terms of their estimates of what they term the "EJ gap," we suggest below that there may be reasons to question that assertion.

We see at least two interpretive challenges with the HCM approach. First, the policy-induced variation in emission changes, as we have noted, is not random noise – it's the point of the program, and so the goal should be to look for systematic differences in the response. Consider a situation where there are only two facilities; faced with C&T, one firm chooses to reduce its pollution while the other does not, with the gain from the first sufficient to make up for the second's choice to not abate. Assuming a common percentage effect for both evades the distributional dynamic; in the appendix, we show why this matters, how variation is a serious issue in this data set, and what difference it might make to consideration that variation at a facility level based on some measure of proximity to a DAC.

We acknowledge that estimating variation rather than a common percentage effect is a challenging econometric proposition which may explain why we and researchers at California's Office of Environmental Health Hazard Assessment (OEHHA) have taken simpler approaches to looking at individual facility-level responses before and after C&T (Cushing et al. 2018; Plummer et al. 2022).[3] In any case, finding that there is relative reduction overall (and then feeding the average reduction into an air model) usefully demonstrates that cap-and-trade is actually capping something – but the question EJ critics have raised is about the variation in facility response to regulations, particularly when those facilities affect different neighborhoods.

The variation issue is particularly important because applying a common percentage estimate means that heterogeneity in the physical quantity of predicted abatement will vary by initial size. If initial pollution is unequally distributed – which both we and HCM highlight (Hernandez-Cortes and Meng 2023; Pastor et al. 2022), applying a common percentage prediction is likely to predict an improvement in the EJ gap whatever variation in emission changes may have actually occurred.

HCM acknowledge the limits of imposing a common percentage effect and so introduce a robustness test in which the C&T differential trend estimates vary by the initial pollution level of the facility. A systematically differential reduction by initial pollution level is of some interest but it does not address the fundamental issue of variable adjustment across facilities by design – nor is this necessarily a good test of the EJ concern since, as it turns out, initial pollution level may be a poor proxy for EJ impact *per se*.

In this dataset, for example, having multiple nearby facilities may be a strong correlate for EJ communities but such a correlation is not as clear when it comes to the magnitude of pollution at individual facilities (Pastor et al. 2022).[4] An alternative approach, taken by Sheriff, interacts cap-and-trade status with the percent of people of color living "downwind" of emitters, providing a clearer test of the EJ concern; that study finds that there were relative improvements for modeled toxic emissions from C&T polluters in communities of color in California (Sheriff 2023).[5]

The HCM article does not utilize such a direct interaction strategy and might best be thought of as a policy thought experiment of how a mandate for across-the-board equal percentage reductions in pollution would affect the EJ gap in California. This thought experiment, of enormous importance to designing environmentally just approaches to decarbonization, is contrary to both the policy design and the heterogenous experience of California communities during the study period (see the appendix and Plummer et al. 2022).

Do HCM Use the Right Data?

While we believe that imposing a common trend is analytically problematic, we explore several other data issues that raise concerns about the accuracy of the analysis. The first is which facilities fall under C&T. As HCM note in a footnote, they found what they believed to be an anomaly in the data that they downloaded from the state's Pollution Mapping Tool, a dataset that links information on greenhouse gas emission with co-pollutants.[6] That anomaly: thirty nine facilities "switched" status from regulated to unregulated, or vice versa, in 2016-2017. Since the authors wish to ascribe time-invariant status (partly to be able to estimate pre-policy trends), they deal with this by assuming that the previous status in their data was accurate and so they reassign the subsequent data to the prior status, and then later test for robustness by dropping those facilities.[7]

The appearance of switching seems to be partly an artifact of their data assembly: according to the authors, they had a dataset downloaded that went up to 2015, and then added 2016 and later 2017 data to it in steps (Hernandez-Cortes and Meng 2022:10). It was in the last step that they apparently discovered the issue and applied the "fix" noted above. We instead downloaded a full dataset as of 2017 – which did not include any switches in regulatory status. We also consulted the California Air Resources Board (CARB) and the state's Office of Environmental Hazards and Health Assessment (OEHHA) to maintain consistency with the C&T tags used in the CARB/OEHHA analysis (Plummer et al. 2022).[8]

In a productive exchange of comments about this issue, HCM note that "Why C&T regulation, or treatment, status has changed across versions of CARB's dataset is worth looking into, in consultation with CARB data managers. However, we note that misassignment of treatment status may not be an issue if such errors occur at random." (Hernandez-Cortes and Meng 2022:10) As noted, we did look into it with CARB data managers. After an extended back-and-forth with the actual data providers, we adopted the C&T status at the end of the post-policy period that OEHHA used in its own study, particularly since that squares with how CARB has issued its official data in the years since.[9]

Of course, this would be of little concern if coming into agreement with OEHHA and CARB made little difference to the results. That is not the case. To check on the impact of using the verified status, we started by following the data restrictions imposed by HCM (no refineries, no electricity generators, and no larger emitters) and utilizing their C&T tags; doing that, we were able to exactly reproduce their Table 1 results and these are shown in the first column of our own Table 1 below.[10] Note that HCM estimate a sharply downward post-policy trend ($\kappa_1^p + \kappa_2^p$) for all pollutants, with nearly all estimates denoted as significant. Feeding the estimates from those equations into an air model is likely to yield impressive improvements overall, with size effects particularly large where pollution was highest to begin with.

In the second column of Table 1, we utilize the C&T tags that came from the 2017 Pollution Mapping Tool and discussion and alignment with CARB and OEHHA. First, note that the number of facilities covered is reduced by 34 when we used the corrected (or verified) C&T tags. The reason: HCM include in their control group facilities that were designated in the data with a "blank" rather than a "Yes" or "No." These are facilities that have emissions in the data from 2008 to 2010 but not thereafter. While that count includes a few facilities that actually shut down before cap-and-trade was implemented, many others are biomass or other facilities that were initially required to report to the state's GHG inventory but were then allowed to stop reporting before the C&T program began (even though nearly all kept operating and emitting co-pollutants). They were never subject to cap-and-trade regulation and they disappear from the database before the policy, and hence we are not convinced that they should be assigned to the control group for considering policy impacts (although including them as part of the control group would not substantively alter what we present).[11]

What do we find when we use the verified C&T tags for the facilities? Looking at second column, one can see that we obtain very different results: while the pre-policy trends are similar, the policy break trends fall and the post-policy trends are just barely negative and always insignificant, with *p*-values rarely below 0.20. For GHG, the post-policy break falls by half, while remaining significant. The drop in the post-policy break is about 1/3 for the co-pollutants, and none retain statistical significance. The post-policy trend (the net of the pre-policy trend and the post-policy break) is always close to zero. For example, a post-trend estimate of -.111 for GHGs in the HCM regression fell in absolute terms to -.015 while the post-trend estimate for PM2.5 shifts from -.039 to -0.008. Other pollutants (PM2.5, NOX, and SOX) had similar patterns.

Table 1. Coefficient Estimates from Five Specifications

Table 1. Coefficient Estimates from Five Specifications

		First Estimate (HCM, Table 1)	Second Estimate (verified, cap-and-trade tags)	Third Estimate (verified & within-unit)	Fourth Estimate (verified, within-unit, CEIDARS data)	Fifth Estimate (HCM C&T, asinh specification, full sample)
TotalGHG	*pre-policy trend*	0.187 ***	0.146 ***	0.124 ***		0.246 ***
		(0.052)	(0.036)	(0.032)		(0.060)
	post-policy break	-0.297 ***	-0.162 ***	-0.124 *		-0.239 ***
		(0.077)	(0.058)	(0.069)		(0.070)
	post-policy trend	-0.111 ***	-0.015	0.000		0.007
		(0.036)	(0.038)	(0.051)		(0.040)
	# of observations	*2,054*	*1,957*	*1,208*		*5,528*
	# of facilities	*316*	*282*	*135*		*758*
	# of C&T facilities	*106*	*112*	*91*		*329*
PM2.5	*pre-policy trend*	0.058 #	0.061	0.052		-0.014
		(0.043)	(0.047)	(0.041)		(0.030)
	post-policy break	-0.097 *	-0.069 #	-0.049		0.009
		(0.048)	(0.053)	(0.045)		(0.049)
	post-policy trend	-0.039 **	-0.008	0.003		-0.005
		(0.018)	(0.018)	(0.003)		(0.025)
	# of observations	*1,968*	*1,869*	*1,177*		*5,244*
	# of facilities	*302*	*268*	*130*		*728*
	# of C&T facilities	*104*	*108*	*89*		*327*
PM10	*pre-policy trend*	0.083 **	0.086 #	0.081 **	0.025	-0.017
		(0.033)	(0.035)	(0.032)	(0.021)	(0.031)
	post-policy break	-0.117 ***	-0.084 *	-0.072 *	-0.018	0.017
		(0.040)	(0.043)	(0.041)	(0.039)	(0.050)
	post-policy trend	-0.034 *	0.002	0.009	0.007	0.000
		(0.018)	(0.016)	(0.021)	(0.024)	(0.026)
	# of observations	*1,968*	*1,869*	*1,177*	*2,508*	*5,244*
	# of facilities	*302*	*268*	*130*	*257*	*728*
	# of C&T facilities	*104*	*108*	*89*	*95*	*327*
NOX	*pre-policy trend*	0.075 *	0.066 #	0.027	0.010	-0.033
		(0.039)	(0.044)	(0.037)	(0.025)	(0.034)
	post-policy break	-0.104 **	-0.057	0.035	-0.003	0.036
		(0.050)	(0.051)	(0.056)	(0.035)	(0.043)
	post-policy trend	-0.029 #	0.009	0.062 *	0.007	0.003
		(0.019)	(0.019)	(0.037)	(0.019)	(0.019)
	# of observations	*1,970*	*1,871*	*1,177*	*2,492*	*5,247*
	# of facilities	*303*	*269*	*130*	*256*	*729*
	# of C&T facilities	*104*	*108*	*89*	*95*	*327*
SOX	*pre-policy trend*	0.006	0.005	0.004	0.013	-0.019
		(0.035)	(0.032)	(0.037)	(0.021)	(0.031)
	post-policy break	-0.037	-0.030	-0.027	-0.022	0.024
		(0.043)	(0.045)	(0.057)	(0.031)	(0.040)
	post-policy trend	-0.031 #	-0.025	-0.023	-0.010	0.005
		(0.019)	(0.022)	(0.027)	(0.026)	(0.017)
	# of observations	*1,965*	*1,866*	*1,174*	*2,455*	*5,229*
	# of facilities	*303*	*269*	*130*	*252*	*729*
	# of C&T facilities	*104*	*108*	*89*	*92*	*327*

*** significant at the .01 level; ** significant at the .05 level; * significant at the .10 level; # significant at the .20 level

The importance of HCM's inclusion as part of the control group those facilities that stopped reporting as of 2011 (and hence were recorded as "blanks" in the data) raises another consideration: whether a facility had observations both before and after C&T. In our next regression, reported in Column 3 of Table1, we specifically required that any facility included in the HCM subsample have at least one observation in 2008-2011, one in 2012 on eve of the policy (in order to estimate a final pre-policy reference point), and one in the post-policy period of 2013-2107. This is not a requirement of the HCM estimating approach. While they list 316 facilities in Table 1, only 135 of those (about 43 percent) meet the "within-unit" criteria suggested above. In short, the majority of the facilities from which they derive their estimates are not providing data on both sides of the policy break.

As noted, 34 of those facilities are the "blank" reporters that do not make this "within-unit" restriction because they stopped reporting emissions since they never came into the trading system; thus, they contribute to the pre-policy trend (in the sense that HCM's estimates of C&T facility trends are benchmarked against them) but they play no role in the subsequent baseline against which one can measure differential C&T post-policy trends.[12] However, most of the decline from, say, 282 facilities for the GHG series in the second panel of Table 1 to the 135 facilities in the third panel requiring within-unit observations is because of 97 facilities that began reporting in 2012 (the last year of the pre-policy period, hence contributing little to the pre-policy trend) and 49 that began reporting in 2013 (in which case they contribute only to the post-policy differential trend estimate).[13]

HCM have suggested that estimating trends in such an unbalanced panel might not be problematic if entry and exit was random (Hernandez-Cortes and Meng 2022:9).[14] However, when 34 facilities fall out of the size and sector-constrained database in 2011, 97 facilities appear in 2012, and another 49 start reporting only in 2013, the pattern seems decidedly non-random.[15] There were shifts in the reporting requirements to GHG mandatory reporting that led to a large number of facilities appearing in the data for the first time in 2012, for example, even though they had been in operation well before then.[16] In short, they have abruptly appeared in the Pollution Mapping Tool data but not in the world.

In any case, at least one reasonable robustness test would be to see what happens when one limits the sample to facilities that have observations on both sides of the trend break, which limits identification to policy-induced within-unit changes in facilities that continue to report.[17] When we do that and use the C&T tags provided by CARB in the third column of Table 1, the estimated post-trends for GHGs and PM2.5 are close to zero. Since these are relative trends, what this means is that the cap-and-trade sector in this regression was doing little better at reducing both GHGs and co-pollutants than the control group.

While this finding of essentially no difference in the post-policy trend might seem odd given that the primary intention of the C&T policy is to reduce GHG emissions more than might occur otherwise, it is likely that the sector and facility size matter; in this subsample of smaller facilities, we may have more reporters that were likely to forgo expensive investments to reduce local GHG emissions and instead purchase allowances. Since one hopes that the cap did have some bite on emissions overall, this suggests the possibility that the HCM sector- and size-constrained subsample – selected to estimate a "pure" effect – may not be representative of all facilities.

Since one might be concerned about the reduction is the number of facilities considered, there is yet another reasonable approach to this issue of the sensitivity of the analysis to requiring within-unit reporting. The Pollution Mapping Tool from which HCM draw their data combines GHGs from one reporting source and co-pollutants from another, specifically the California Emissions Inventory Development and Reporting System (CEIDARS). We obtained data from CEIDARS for the years 2008 to 2017 and matched it in with the facilities in the Mapping Tool.[18] We were able to link in nearly all the facilities with one set of exceptions: the Pollution Mapping Tool includes a select number of oil and gas emitters that are really reporting entities for satellite emitters at multiple, highly dispersed locations.[19]

The advantage of linking to the underlining CEIDARS data is that we get a more precise and longer duration accounting of the pattern for three co-pollutants, PM10, NOX, and SOX, for their both the pre-policy and policy periods.[20] In this more complete dataset, the pattern of entry and exit from the data is less pervasive and more plausibly random: of the 256 facilities in a PM10 regression for the HCM size- and sector-constrained sample, 231 start in 2008 and end in 2017, and the rest of the series begin and end in other years without the bunching of exits in 2010 and entries in 2012 and 2013 that are seen in the HCM regressions.

The results of this approach are shown in the fourth column of Table 1. First, note that the number of facilities included now is quite close to those in the second column – that is, we have quite good coverage for the co-pollutants even when we require within-unit observations from before and after the policy was implemented. The results again suggest a flat post-policy trend relative to the control group (although to be clear, none of the coefficients for any relative trend affect attains significance in size and sector restricted subsample used by HCM).[21]

Finally, recall that HCM suggest that the results for the full sample are generally the same as for their sector- and size-constrained sample. While we are able to exactly reproduce HCM's results for their restricted sample, HCM do not present a table with the stage 1 results (prior to the air model) for the full sample. In the fifth column of our Table 1, we show the results from regressions in which we apply HCM's analytical approach over the entire sample, that is, we use the inverse hyperbolic sine (*asinh*) specification and the HCM (non-verified) C&T tags which includes the "blanks" that we suggested should not really be in the control group. The post-policy estimates are all close to zero which is, as we have noted, not likely to generate a reduction in the size of the EJ gap.[22]

How can these results be consistent with the assertion in their paper that their basic findings hold across the whole sample? As it turns out, Table S12 in HCM's published paper – which is meant to show the results of a process of applying the asinh specification to generate predictions over the whole sample and then running those through the air model to generate an estimate of the EJ gap – has coefficients and standard errors for PM2.5, PM10, and NOX that are exactly identical to the coefficients and standard errors for those pollutants in Table 1 of the 2020 working version of their paper which relied on a log (rather than an inverse hyperbolic sine) transformation of the dependent. In the appendix, we note that there are reasons why dropping observations values that are zero (which is what a log specification does) is likely to generate different results.

The basic conclusions of this data exploration are straightforward. First, it matters which set of cap-and-trade tags one uses; we believe there are good reasons to use the tags that were verified by CARB and OEEHA. Second, while an unbalanced panel may pose no problems for some analyses under particular conditions of random entry and exit, looking at before-and-after effects in which the majority of facilities in question do not appear both before and after the policy change seems problematic, particularly because the entry and exit timing into the data is not random and does not in fact generally capture the actual entrance or exit of facilities into the world. Restricting the sample to include observations on both sides of the policy break or supplementing the sample with a fuller range of data from the original CEIDARS source for the co-pollutants yields different results. Finally, whether or not the HCM results on the EJ Gap hold across the whole sample is not a settled question given the regression results we obtain and the evidence HCM have presented.

Conclusion

We conclude this comment as we began: we appreciate that HCM have advanced the field with air modeling and we concur with their policy conclusion that market-based approaches will not necessarily improve (or worsen) EJ gaps, and so EJ-specific policies must be put in place (Hernandez-Cortes and Meng 2023:15). However, the bulk of HCM's empirical conclusions depend, at least in part, on whether the constant percentage estimate of the C&T effect fed into their air model is relevant to the policy question at hand, whether the estimates may be impacted by any non-random data irregularities, and whether the subsample on which they focus is illustrative or anomalous.

Assuming a common percentage effect does not effectively address the EJ concern about disparities in adjustment effects and, as we note in the appendix, is not a particularly good description of the data. We would suggest that both future analysis of the EJ implications and future policy design to incorporate EJ in carbon regulation should directly address the variation in response among C&T facilities. In that regard, concerned policymakers might benefit from applying HCM's sophisticated plume air modeling to actual emissions rather than common percentage predicted emissions to better assess the variation in policy-induced local-pollution impact across EJ communities on California.[23]

Our work further suggests that even by the standards of a common percentage approach – which we think addresses the cap and not the trade portion of the program – the results obtained by HCM are highly sensitive to a set of data decisions with regard to which facilities are actually regulated by cap-and-trade, whether one should focus on within-unit analysis, and whether one should deal with dataset issues having to do with non-random entry and exit into the reporting system by including the full set of available co-pollutant data.

While we have concerns about cap-and-trade, we should be clear that we are not opponents of carbon pricing and have actually argued that a few environmental justice safeguards could help insure a more favorable reception to such efforts (Boyce, Ash, and Ranalli 2023; Boyce and Pastor 2019). It is also quite possible that the data patterns will reveal an improvement in EJ outcomes as the program continues over time. However, the key point of this comment is that the jury is still out on the impact of market strategies on environmental justice. Further research is needed and the advances offered by HCM will be an important part of that effort.

Endnotes

[1] The *asinh* $= \ln\left(x + \sqrt{x^2 + 1}\right)$ refers to an inverse hyperbolic sine, a function that performs much like taking a natural log, i.e., with differences usually translating into percentage changes, but allows for the retention of observations where the value to be transformed is zero or negative. As we note in the appendix, while retaining zero's has a certain appeal, it also creates estimation problems in this particular dataset.

[2] Whether the trend break, κ_2^p, is a meaningful signal also depends on the reliability of the pre-trend estimate; in in the appendix, we discuss why the pre-trend estimates HCM offer may be problematic, particularly for GHGs.

[3] In a recent effort, Currie et al. do try to get at this issue by employing a "triple difference" regression approach to looking at the impact of the Clean Air Act on PM2.5; the findings suggest that within-county improvements were less for African Americans than for non-Hispanic whites although the overall story is that their version of the EJ gap shrank because areas of the country with a higher share of Black residents were more likely to see improvements due to regulations (Currie, Voorheis, and Walker 2023). The authors also utilize a quantile regression approach to model differential impacts on different points on the national pollution distribution. In general, this is a more convincing approach to estimating variation – and the exploration of variability is both useful and admirable given that the policy (unlike cap-and-trade) is not intended to create variation or trade-offs but rather to bring every area out of a "nonattainment" status.

[4] For example, for facilities HCM designate as cap-and-trade for which we were able to link average CalEnviroScreen scores within five miles, the correlation between HCM's preferred size metric (average annual metric tons of GHGs) and that CalEnviroScreen measure is .0017, with a significance level of .9872, essentially a finding of no relationship.

[5] Sheriff (2023) connects C&T and non-C&T facilities with air-modeled data on toxics taken from the U.S. EPA's Toxic Release Inventory (TRI); for those familiar, Sheriff specifically uses the Risk-Screening Environmental Indicators (RSEI) model which we have also used in other work (see, for example, Ash et al. (2012)). Whether this also applies to other co-pollutants, like PM2.5, is an open empirical question but Sheriff does point to an approach that could use a difference-in-difference regression and allow variation in this setting a la Currie et al. (Currie et al. 2023).

[6] One of the co-authors worked on an early report that linked the GHG and co-pollutant data before the state took on this task (Cushing et al. 2016); it was the demonstration that it was both possible and of public interest that prompted the Air Resources Board to do the data assembly for broader use that resulting in the Pollution Mapping Tool. This familiarity with the underlying data construction explains in part our facility in relinking the reporting facilities to the underlying co-pollutant data, something that will become important later in this analysis.

[7] If HCM had tested what would have happened if they had not reassigned facilities to what they thought was the previous C&T status (rather than testing for robustness only by dropping those facilities), they would have gone part of the way to addressing some of the issues we raise.

[8] To give the reader a flavor of the size of the issue, HCM's GHG regression has 316 facilities in the subsample (that are not singletons and so will enter the regression). Of those, 22 were switched by HCM from their new reported values to the older values. 3 of those switches were valid – the facilities had actually stopped reporting and the last reported values were accurate. 19 of the 22 switches did not agree with what was in the consistent 2020 downloaded database and likely should not have been switched. All of those facilities were checked with OEHHA facility-by-facility and "verified." Upon the direction of OEHHA (and to be consistent with their own study), another four in the subsample were re-tagged as a "Yes"; those changes were eventually implemented by CARB for the relevant years in subsequent versions of the Pollution Mapping Tool, suggesting that these were appropriate corrections to make.

[9] Along with OEHHA, in our own analysis, we made a series of additional corrections to the data that included dropping facilities where either a large increase or a large decrease in pollutants had little to do with cap-and-

trade (such as a firm that shut down operations after being targeted for lead emissions or facilities that significantly reconfigured their reporting for other reasons – such as acquiring new plants from other firms) or where there were known data errors. We did not apply these extra data corrections when trying to replicate the HCM results here although there may a good reason to do so if accuracy rather than replication were the goal; we discuss in the appendix how the failure to apply such corrections can also affect coefficient estimates.

[10] Aside from the additional corrections noted in the previous footnote, there are slight differences in the datasets used by HCM and our team, partly because our data which was updated throughout as of 2017 (rather than added to the 2008-2015 data in two separate steps) had some backward corrections of emissions levels. To reproduce the HCM results, we used their data and simply applied both their C&T tags and the C&T tags provided by state agencies. Both research teams exchanged data and code as we were seeking to resolve some of the inconsistencies; both teams wanted to be sure that the differences in results reflected differences in approach, and we thank HCM for the collegial spirit that has characterized our exchanges.

[11] Of the 34 facilities in the HCM subsample that stop reporting GHG emissions to the Pollution Mapping Tool data, 30 can be linked to a co-pollutant database we describe later, with 27 of those having co-pollutant observations, and 25 of those have co-pollutant data that stretches to 2015 or 2017. In short, these facilities stopped reporting their greenhouse gas emissions but the overwhelming majority continued to operate; they were tagged as "blank" in the data because they were temporarily required to report as CARB was setting up its reporting in 2008-2010 but were never subject to potential regulation under cap-and-trade. In our view, they are not really part of the control group and if they were, one should definitely not include them as though they actually stopped operation in 2010 – when the overwhelming majority did not. As it turns out, bringing them back in as "No's" – which we think is inappropriate – would not alter the pattern of results we discuss.

[12] These facilities are actually dropped out in the first round in which we applied the CARB-verified C&T tags because they were coded in the 2017 downloaded data as "blank" – because they were never in the system – but counted by HCM as "No" for the pre-policy period.

[13] One additional facility also drops out because it has no observations for 2012. As we will see when we look at the actual CEIDARS data, these facilities did not all start up in those years; rather, it was the first time they began reporting to the inventory. As such, the HCM regression is not really capturing their pre-policy trend.

[14] In an initial set of exchanges with HCM, the authors suggested that they required two observations before and after the imposition of policy change to be able to estimate trends – but as it turns out, they did not. A growing literature underlines the importance of clean controls and panel balance (Cengiz et al. 2019).

[15] The problem seems to be a bit more pronounced in the size and sector-constrained subsample HCM use in their regression. For the 758 facilities that have at least two years of observations (and so would make it into an HCM-style regression with no restrictions on before and after observations – since singletons drop out in the regression exercise itself), 173 would fall out for starting in 2012 or later which amounts to less than a quarter of the sample versus the fall-out of nearly half the facilities for a late start in the subsample HCM think is most appropriate for the testing of effects.

[16] See the 2008-2012 Emission Summary at https://ww2.arb.ca.gov/mrr-data The new facilities were smaller and thus more likely to impact the HCM regression sample as hinted at in the previous footnote. We thank Danny Cullenward for explaining to us the shift in reporting requirements.

[17] Shutdowns and entry of facilities are potentially important responses to the cap-and-trade program. The data do not permit HCM or us to systematically identify exiters or entrants in response to the program itself; as we note in the appendix, when we do investigate the history of those facilities that shut down, the reasons seem to have little to do with cap-and-trade. In the replication portion of this exercise, we follow HCM and retain the exiters as though this was a consequence of the program; in the appendix, we show how that approach can severely bias the estimates of trends.

[18] We originally did this because the Pollution Mapping Tool reports rounded estimates for the co-pollutants and such rounding led to zero observations on co-pollutants (unlike on GHGs) when there were really were positive emissions; as such, we wanted to check on sensitivity to improved data accuracy in our work in (Pastor et al.,

2022). As it turns out, that actually had little impact on our analysis but we realized that it provided a way to cover more years of data for this replication.

[19] As we discuss in our own work, a share of the oil and gas facilities in the Pollution Mapping Tool are central reporting entities for multiple emitters that are often a long distance away. In our analysis, we obtained all the underlying locations and co-pollutant reports so we could determine the impact on neighborhoods. However, we did that only for cap-and-trade rather than control facilities and the additional level of work to link CEIDARS data for all reporters, including non-C&T reporters, seemed excessive for this replication and so we simply dropped that limited number of reporters that essentially collect and aggregate data from multiple satellites; in the HCM subsample, that amounted to ten reporters. However, it should be noted that HCM assume that the geographic location of these select reporting entities given in the Pollution Mapping Tool is where the emissions are actually released – it is not and we think a correction to the underlying locations would be best for the air model since one would likely want the plume to start in the right place.

[20] PM 2.5 was not easily available as it requires additional processing by state authorities.

[21] If we relax our assumption that the "blanks" really are not part of the control group and instead set them as "No" and include the full CEIDARS data for those facilities if it exists on both sides of the policy break, the number of observations and facilities rises slightly but the general pattern of insignificant remains the same.

[22] In the appendix, we also explore the fit of the HCM model using an event study approach. This also suggests that there may be a poor fit between the HCM estimates offered in their Table 1 and the actual post-policy pattern for the whole sample.

[23] OEHHA does employ air modeling in their analysis of cap-and-trade, although their approach is similar to ours in the use of simple comparisons, in their case of quantity versus percentage impacts (Plummer et al. 2022). They find some EJ improvements using that aggregate approach, adding to the picture of mixed results. Interestingly, we suspect that the issue of the EJ effects of cap-and-trade is likely to be of continued interest to academics but of lesser interest to policy makers as the portion of GHGs regulated by market strategies seems to be on the decline in California and running into resistance elsewhere.

References

Ash, Michael, James K. Boyce, Grace Chang, and Helen Scharber. 2012. “Is Environmental Justice Good for White Folks? Industrial Air Toxics Exposure in Urban America.” *Social Science Quarterly* 94(3):1–21. doi: 10.1111/j.1540-6237.2012.00874.x.

Boyce, James K., Michael Ash, and Brent Ranalli. 2023. “Environmental Justice and Carbon Pricing: Can They Be Reconciled?” *Global Challenges* 2200204. doi: 10.1002/gch2.202200204.

Boyce, James K., and Manuel Pastor. 2013. “Clearing the Air: Incorporating Air Quality and Environmental Justice into Climate Policy.” *Climatic Change* 1–14. doi: 10.1007/s10584-013-0832-2.

Boyce, James K., and Manuel Pastor. 2019. “Can Carbon Pricing Address Climate Justice?” *The Nation*, November 7.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. “The Effect of Minimum Wages on Low-Wage Jobs*.” *The Quarterly Journal of Economics* 134(3):1405–54. doi: 10.1093/qje/qjz014.

Currie, Janet, John Voorheis, and Reed Walker. 2023. “What Caused Racial Disparities in Particulate Exposure to Fall? New Evidence from the Clean Air Act and Satellite-Based Measures of Air Quality.” *American Economic Review* 113(1):71–97. doi: 10.1257/aer.20191957.

Cushing, Lara, Dan Blaustein-Rejto, Madeline Wander, Manuel Pastor, James Sadd, Allen Zhu, and Rachel Morello-Frosch. 2018. “Carbon Trading, Co-Pollutants, and Environmental Equity: Evidence from California’s Cap-and-Trade Program (2011–2015).” *PLOS Medicine* 15(7):e1002604. doi: 10.1371/journal.pmed.1002604.

Cushing, Lara, Madeline Wander, Rachel Morello-Frosch, Manuel Pastor Jr, Allen Zhu, and James Sadd. 2016. *A Preliminary Environmental Equity Assessment of California’s Cap-and-Trade Program*.

Duan, Naihua. 1983. “Smearing Estimate: A Nonparametric Retransformation Method.” *Journal of the American Statistical Association* 78(383):605–10. doi: 10.1080/01621459.1983.10478017.

Hernandez-Cortes, Danae, and Kyle C. Meng. 2022. “The Importance of Causality and Pollution Dispersal in Quantifying Pollution Disparity Consequences: Reply to Pastor et al. (2022).”

Hernandez-Cortes, Danae, and Kyle C. Meng. 2023. “Do Environmental Markets Cause Environmental Injustice? Evidence from California’s Carbon Market.” *Journal of Public Economics* 217:104786. doi: 10.1016/j.jpubeco.2022.104786.

Pastor, Manuel, Michael Ash, Lara Cushing, Rachel Morello Frosch, Edward-Michael Muna, and James Sadd. 2022. *Up in the Air: Revisiting Equity Dimensions of California’s Cap-and-Trade System*. Los Angeles, CA: USC Equity Research Institute.

Plummer, Laurel, Amy Budahn, Annie I. Chen, K. Lily Wu, and Álvaro Alvarado. 2022. *Impacts of Greenhouse Gas Emission Limits Within Disadvantaged Communities: Progress Toward Reducing Inequities*. Sacramento, CA: Office of Environmental Health Hazard Assessment.

Sheriff, Glenn. 2023. “California’s GHG Cap and Trade Program and the Equity of Air Toxic Releases.” *Journal of the Association of Environmental and Resource Economists*. doi: 10.1086/725699.

Stavins, Robert N. 1998. *Market-Based Environmental Policies*. *Discussion Paper*. 98–26. Washington, DC: Resources for the Future.

Zwickl, Klara, Simon Sturn, and James K. Boyce. 2021. “Effects of Carbon Mitigation on Co-Pollutants at Industrial Facilities in Europe.” *The Energy Journal* 42(01). doi: 10.5547/01956574.42.5.kzwi.

Appendix. A Few More Data and Analytic Concerns

In this appendix, we explore a few more data and analytic issues. We start by discussing the reliability of HCM's pre-trend estimate, κ_1^p, and post-trend estimate, κ_2^p for GHGs. We then contrast the actual variations in the data and the predictions for co-pollutants that emerge from the HCM specification. Finally, we report on a non-parametric event study which questions the fit of the HCM approach for the whole data set.

The Reliability of the GHG Estimates and the Implications

We decided to investigate the GHG estimates because the coefficient estimates on the GHG series are rather surprising. For example, CM report that "C&T reduced emissions annually between 2012–2017 at an average rate of 9% . . . for GHG" (Hernandez-Cortes and Meng 2023:2), a pace that bears little resemblance to what actually happened to GHG emissions in the cap-and-trade sector and would be a remarkable overachievement given that the cap was only calling for a three percent annual reduction over the relevant time period.

Of course, the estimated reduction is partly dependent on the steepness of the pre-policy trend, and there are some reasons to be cautious about that since there was a shift in reporting requirements for GHGs and PM in 2011 – significant enough that there is a warning, particularly about GHGs, on CARB's website for Mandatory GHG reporting ((see https://ww2.arb.ca.gov/mrr-data). The agency notes that one should not connect the 2008-2010 data to the data from 2011-on – which is why we did not do that in our own analyses (Cushing et al. 2018; Pastor et al. 2022). Of course, data challenges occur all the time and HCM (2022:10) argue that "there is no evidence that these reporting changes were different for C&T regulated and unregulated facilities . . . (and) any common reporting changes is addressed in HCM's equation 1 through year-specific fixed effects."

To see whether that is the case, in Figure 1, we compare the pattern of mean GHGs for non-C&T and C&T facilities over the 2008-2017 period that are in the HCM sector- and size-constrained sample. To insure that the pattern is not determined by facilities exiting and entering but by shifts in reporting, we require that each facility report observations for all years. As can be seen, not much happens to average emissions in the C&T sector over the entire period, while there is a large decline in non-C&T emissions between 2010 and 2011, prior to the advent of the policy in 2013, and then a slow and steady decline.[1]

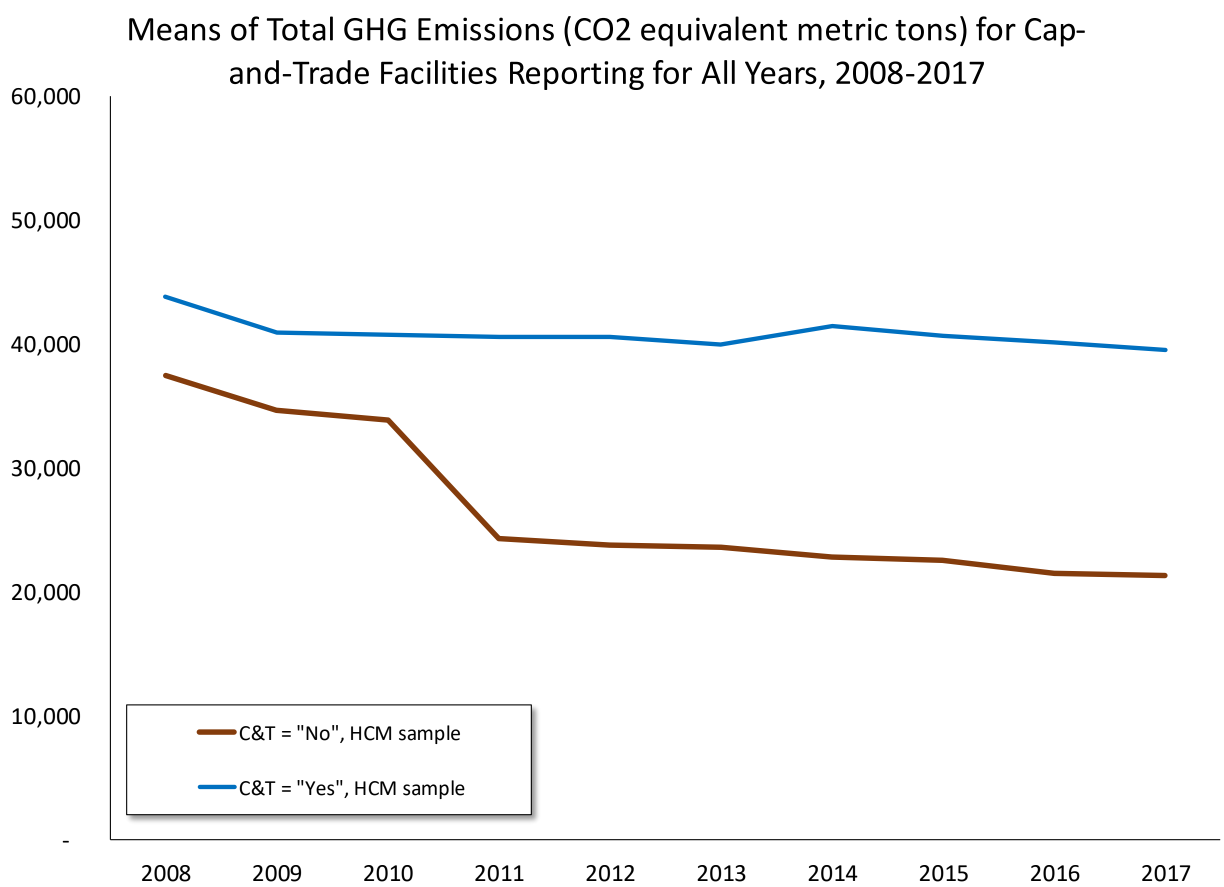
Consider how the failure to account for this apparent shift could lead to overstatement of the pre-policy trend and the post-policy shift. After all, while the pre-trend κ_1^p is a relative estimate, it tends to draw one's attention to the notion of pollution rising up in the C&T sector prior to policy adoption.[2] Of course, the flip side is why relative pollution went down so dramatically in the non-C&T sector: if it was due to

[1] To verify that this is not just the result of the sectoral constraints, we also looked at the pattern for all facilities that have all years of data regardless of sector that meet the size threshold (so that the means are not distorted by large outliers). That pattern is nearly identical.

[2] For example, when we first saw the HCM results, we wondered whether firms were stepping up pollution prior to having their reduction targets set – we should note that we also wondered why there were such steep declines when that didn't occur in the world but we turn to that below.

the differential impact of reporting changes in the non-C&T sector, as suggested by the patterns in the data in Figure 1, then the *relative* pattern for the treatment facilities will be rising sharply in 2011 and 2012.

Figure 1



Why is there a such difference between the C&T and non-C&T sectors in terms of GHG reporting in 2011? It turns out that biomass reporters seem to have be more affected by the shift in reporting requirements that occurred in 2011. For example, if we just concentrate on non-C&T in the HCM sample, and look at the shift in total GHGs between 2010 and 2011, the decline is about 50 percent for the high biomass firms and about 3 percent for the low biomass firms. As it turns out, of the 316 facilities in the HCM regression subsample, 20 of the 21 high-biomass reporters were in the non-C&T category and they are the main reason that there is the sharp drop in non-C&T lines in Figure 1.

Clearly, a commonly shared year fixed effect will not capture this and one needs a dummy variable that can account for the difference in a regression analysis of trends. To see the effect of this on the trend estimates, we created a variable that tagged a facility as high biomass if its associated NAICS code was such that more than half of total GHGs were biomass (with the overwhelming majority of high biomass industries crossing a threshold that was much higher than that). We report on the impact of this dummy variable below.

In any case, this suggests that the pre-policy upward trend in relative GHG emissions in the C&T sector might be overstated. What about the sharp post-policy downward trends? As it turns out, another factor influencing the HCM results on GHG (and to some extent co-pollutants) has to do with facility shutdowns and how those appear in the data. Part of the reason why HCM use an inverse hyperbolic sine function for the dependent variable is so that they can include observations where the reported total is zero.[3] But what does a reading of zero GHGs actually mean if you look at the data carefully?

By direct examination of cases, we found that zero GHG report indicates that the plant shut down but still reported for an additional year or two to the state and so the recorded values are zero. One argument for retaining those observations would be if the shutdowns were actually induced by cap-and-trade – although even then outlier issues might argue for a bit of caution (or at least robustness tests). To see whether that was the case, we examined the cases where there was a zero observation (or, in some cases, near-zero) for total GHGs and then did research on the history of the facility. While we looked at the entire data set, we paid special attention to the nine cases in the HCM size- and sector-constrained sample which would make it into a within-unit regression of the type in the third column in Table 1.

The nine shutdowns in that subsample included a General Mills plant that reported low demand as its cause for stopping operations, two glass manufacturers that faced other environmental mandates, an oil facility where a pipeline broke, a food processor that reported that it shut for market conditions, a beef processor in Brawley, California that reported the same, and a fertilizer terminal that was purchased and seems to have temporarily suspended operation (or at least reporting) shut down during the ownership transition. There is also a timber company that suddenly reported a 99.9 percent decline in emissions to 17 metric tons in the final year of the data; that turns out to have been a misreport for the real figure which was 17,034. Finally, there was Exide Technologies, a battery processing plant which was forced to suspend operations in 2013 and later shut completely because its lead emissions were posing a threat to 100,000 neighboring residents.

In our view, one should exclude those observation-years where the shift to zero GHGs had nothing to do with cap-and-trade, at least to test for the robustness of results.[4] We should stress that none of the regressions included in the main body of the paper tried to make any corrections or exclusions in line with the logic above; recall that there we were striving for replication not redoing. Here, we are trying to highlight what might happen if you were to account for the high-biomass facilities having a different pattern of reporting and for the seeming “shutdowns” that, as we will see, have a big and biased impact on the trend estimates for GHGs.

To follow the action, we trace the impacts the impact of correcting for the biomass reporting and excluding the zero observations in a step-by-step approach presented in Table 2 and a series of accompanying Figures. In the first column of Table 2, we offer the basic HCM results that are in their Table 1 and which were presented in the body of the paper. The visual of this is in Figure 2 where the

---

[3] Relying on the asinh transformation may make sense with the co-pollutants: the Pollution Mapping Tool uses a rounded calculation of various co-pollutants and so using the inverse hyperbolic sine can retain low values that are positive but have been rounded down to zero (the other alternative is to actually go to the original CEIDARS values which have several more digits of accuracy and that is what we did in the fourth regression in Table 1.

[4] We also excluded Excide entirely and dropped the one year of a misreport for the timber company.

solid lines capture what we reported before: the mean of the observations for which we have all years of data in HCM sample categorized by whether the data is from C&T or non-C&T facilities. The dotted line is the mean of the predicted pattern for C&T and non-C&T facilities as indicated by HCM's coefficients.[5]

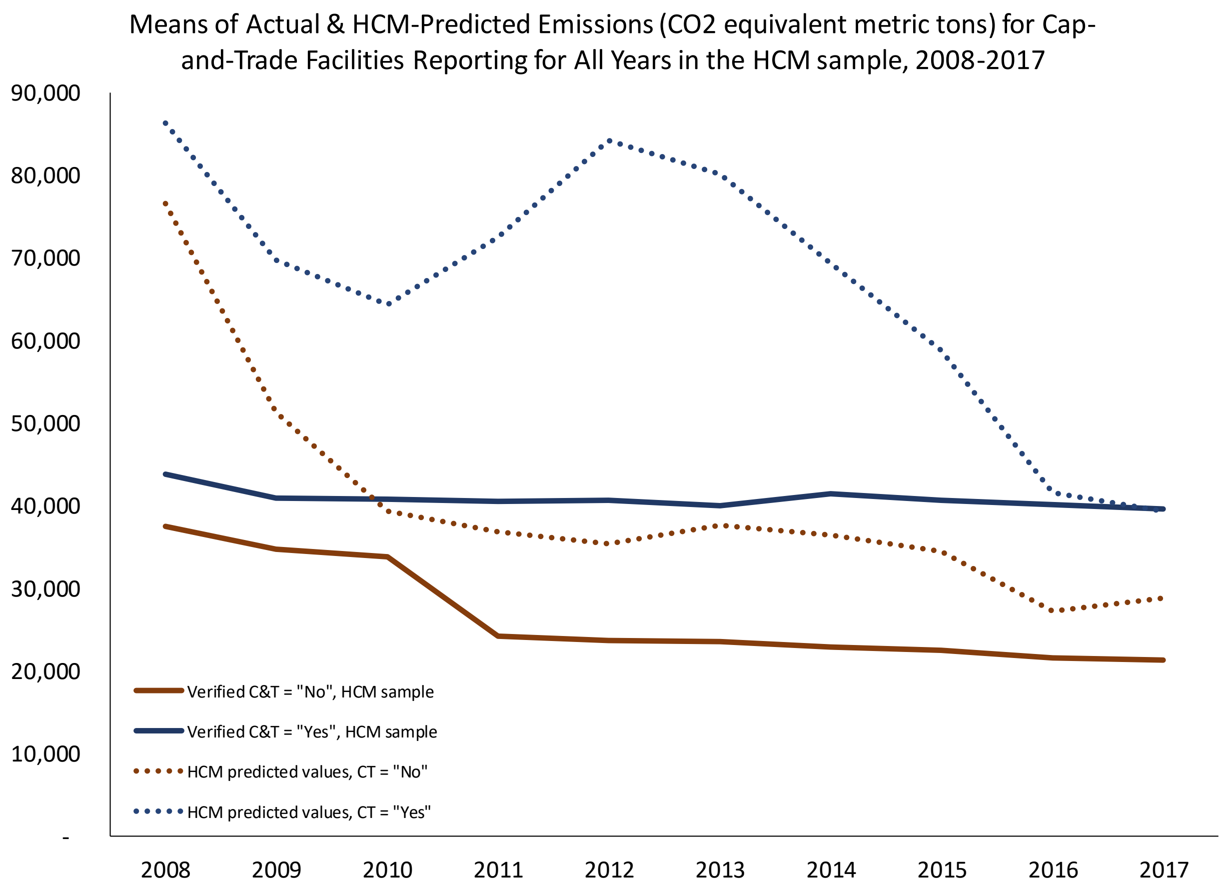
Table 2. GHG Coefficient Estimates from Four Specifications

	First Estimate (HCM, Table 1)		Second Estimate (verified C&T & within-unit observations)		Third Estimate (verified, within-unit, and high-biomass dummy)		Fourth Estimate (verified, within-unit, high-biomass dummy, excluding certain shutdowns)	
pre-policy trend	0.187	***	0.124	***	0.100	***	0.073	***
	(0.052)		(0.032)		(0.032)		(0.026)	
post-policy break	-0.297	***	-0.124	***	-0.099	#	-0.065	**
	(0.077)		(0.069)		(0.062)		(0.027)	
post-policy trend	-0.111	***	0.000		0.001		0.008	
	(0.036)		(0.051)		(0.051)		(0.013)	
High Biomass (2008-2010)					0.299		0.455	
					(0.169)	*	(0.118)	***
# of observations	*2,054*		*1,208*		*1,208*		*1,189*	
# of facilities	*316*		*135*		*135*		*134*	
# of C&T facilities	*106*		*91*		*91*		*90*	

*** significant at the .01 level; ** significant at the .05 level; * significant at the .10 level; # significant at the .20 level

[5] One wrinkle here is that for this, we are necessarily using the HCM tags rather than applying their coefficients to the verified tags; this seems to be a truer approach to presenting the pattern than they are asserting is the case.

Figure 2



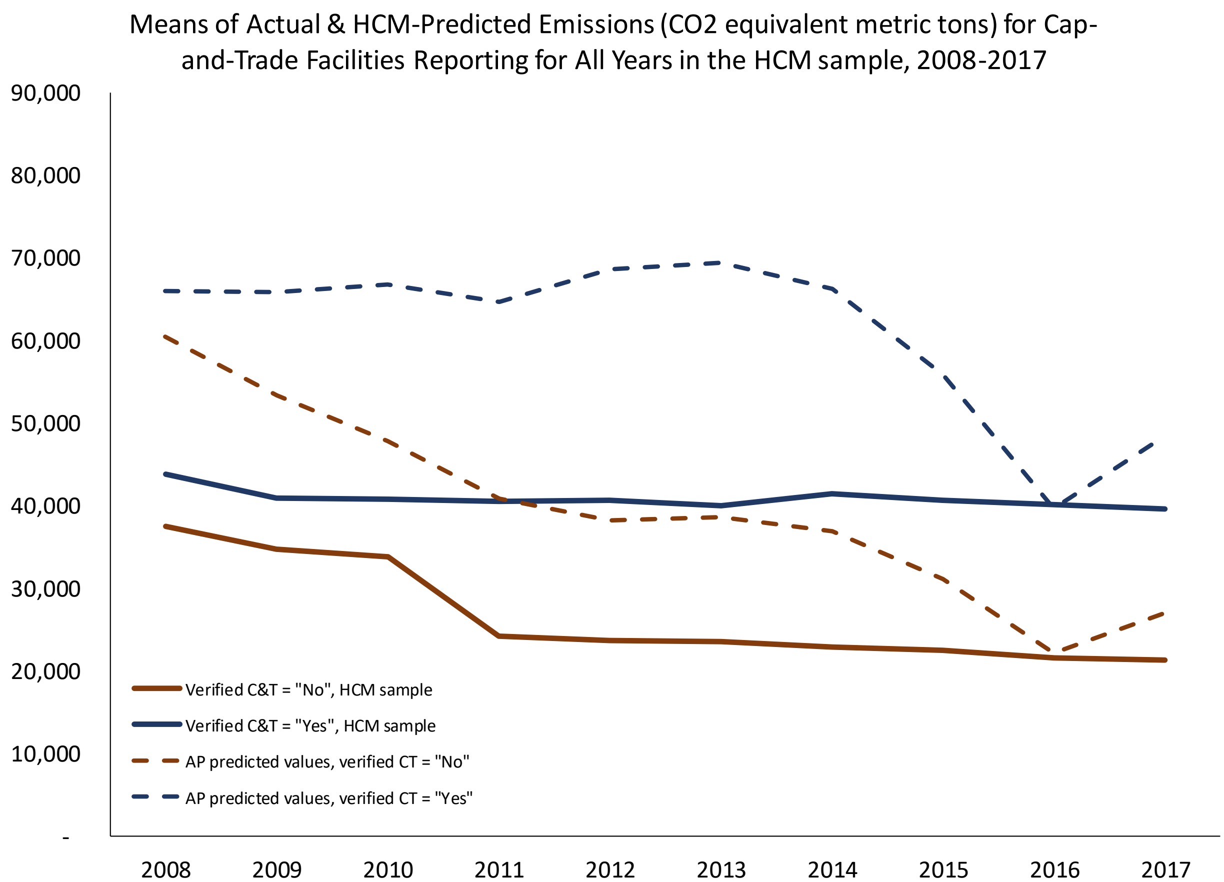
There are two things to notice in Figure 2. The first is perhaps most immediately striking: the predicted values for both C&T and non-C&T facilities are much higher than the actual values. While this might seem surprising, consider that the predicted value that HCM generate is $\widehat{Y_{it}} = \sinh(\widehat{y_{it}}) \cdot e^{\frac{MSE}{2}}$ where $\widehat{y_{it}}$ incorporates the facility fixed effect estimate, the year fixed effect estimate, and the estimated pre-trends and post-trends based on the regulatory-status of the facility from Equation 1, with the expanded value then multiplied by the term involving mean square error (MSE) to account for the concavity of the asinh transformation, as in HCM (2023).[6] Because the HCM regression for GHG has a particularly poor fit, the large MSE drives up the predicted value and the distance between that the actual value.

The second thing to note is the widening gap between C&T and non-C&T facilities in the HCM estimates for the pre-policy years – which would yield a very high estimate of the pre-policy relative trend, κ_1. Note that this overestimated pre-trend is then followed by a very steep decline represented by HCM's C&T estimated trend post-policy, $\kappa_1 + \kappa_2$. Finally, note that the HCM estimates for both C&T and non-C&T facilities differ sharply from the actual pattern indicated by the solid lines capturing the means.

[6] While we adopt HCM's approach to converting the predicted asinh estimates to predicted estimates, a likely superior strategy would be to adopt the Duan smearing estimates which would do better in this sample and result in fewer negative predicted values (Duan 1983).

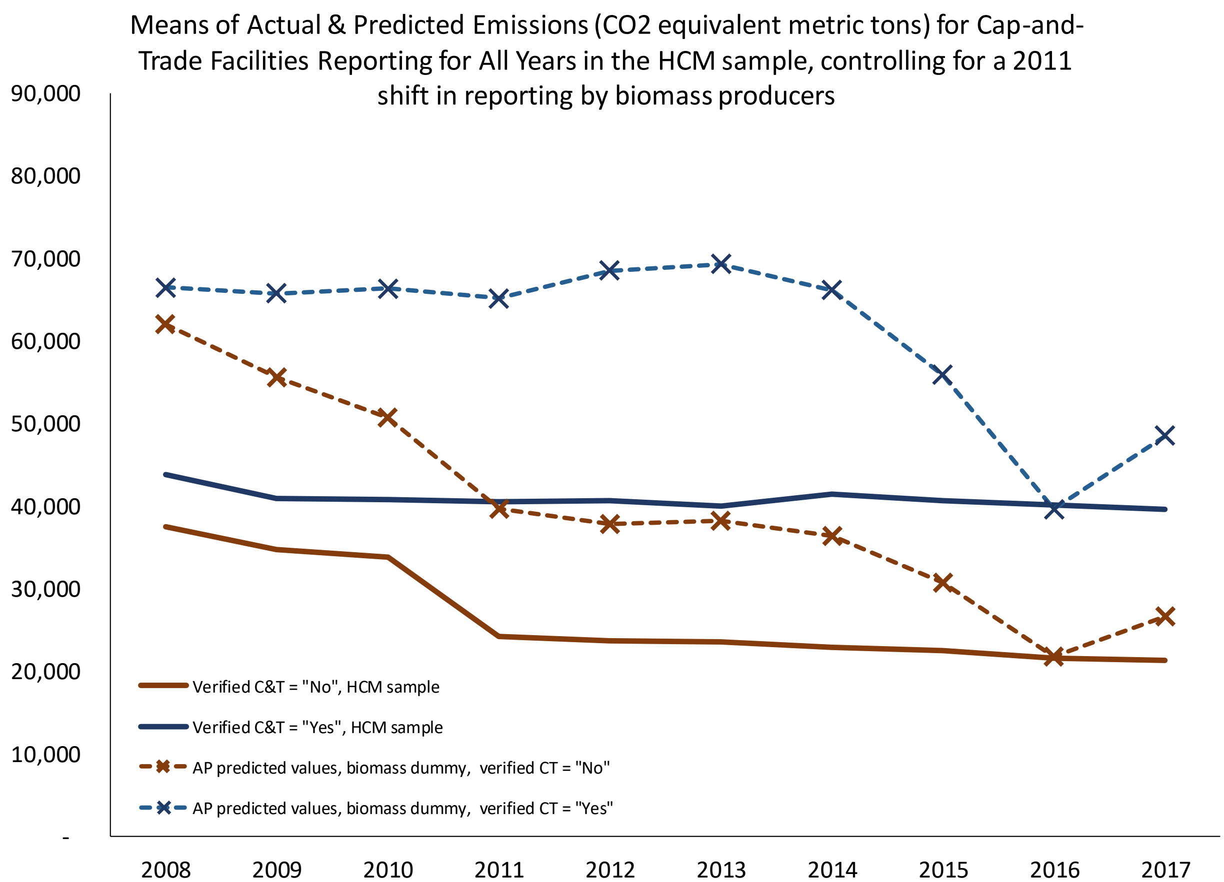
Our approach to estimating what might be a more reliable set of trends proceeded in three steps. First, we shifted to what we think is a more approach regression, that is, one that uses the verified C&T tags and requires that any facility have observations both before and after the policy shift. This is the same as the regression in the third column in Table 1 and it is reproduced in the second column of Table 2. The resulting predicted values are now depicted in Figure 3. As can be seen, this is a better fit with the data; it is closer to the means and patterns of the actual values.

Figure 3



Next, we introduce a dummy variable for 2008-2010 if the facility was in a high biomass industry to capture the reporting shift, and we then applied that to the within-unit estimation. The statistical results are in the third column of Table 2 and the predicted values are shown in Figure 4. Once again, the solid lines are the actual data for facilities that report for all years; the dashed lines with the X's are estimates from those facilities from a regression that includes the high biomass dummy. As can be seen, the estimated pre-policy trend for both C&T and non-C&T facilities is now closer to the actual pattern with a sharp king in 2011; as might be expected and can be gleaned from Table 2, the estimated relative divergence between C&T and non-C&T predicted GHGs is now was lessened to +.10 year from HCM's rather remarkable +.19.

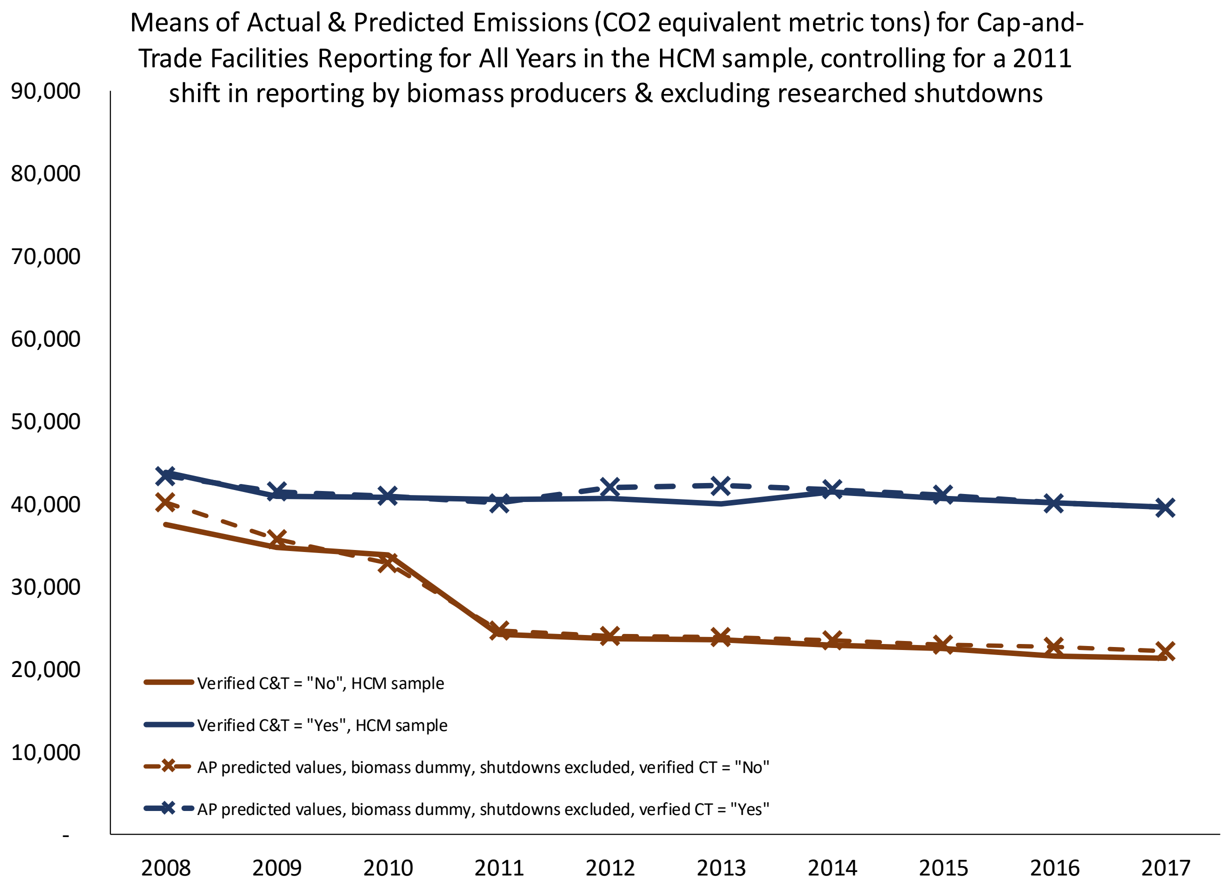
Figure 4



Finally, we ran a regression in which we retained the high-biomass dummy and dropped the observation-year combinations that reflected shutdowns due to non-C&T factors, the Excide lead poisoning shutdown, and the single year of misreported data for the timber company. We acknowledge that there may be some valid theoretical reasons to include shutdowns in an analysis of cap-and-trade – although we are quite sure from our research of the facility histories in the restricted sample that these particular shutdowns and misreports have little to do with cap-and-trade. That, however, is not our main point: we are simply trying to address the empirical issue of how these observations in the data can introduce potential bias in the estimation strategy, and so should at least be considered as part of robustness test.

Figure 5 shows the result of that last step in our analysis. Once again, the solid lines are the actual data for C&T and non-C&T facilities that report GHG emissions for all years; the dashed lines with the X's come from a regression that included both the high-biomass dummy and the elimination of the observation-years discussed above. Note from Table 2 that we have lost relatively few observations and only one facility (Excide). Note further that the fit between the actual means and these predicted values is quite tight both in terms of the pattern and the level. Furthermore, the post-policy relative trend is now .008 or basically flat (the two lines for C&T and non-C&T are basically parallel) as can be discerned from the figure.

Figure 5



To assuage any concerns that this treatment of shutdowns leads to some bias with regard to estimating cap-and-trade impacts, we should note that roughly the same number of C&T and non-C&T facilities are affected by the exclusion of some observation-years. But if both categories are affected, why does the estimated relative divergence shrink so much? Consider two facilities, both of which are trending flat from 2013 to 2016, with one posting 50,000 metric tons per year and the other posting 25,000 per year – and suppose that both fall to zero in 2017. Now imagining fitting a trend line to each of those: the one that was trending higher at 50,000 would have a steeper estimated decline over the period. Now recall that the non-C&T facilities generally trend at a lower emission level than the C&T facilities; when an outlier of zero at the end of a facility's time series due to a shutdown (that should have been excluded) tugs a regression line down, the steeper impact is on the C&T facility and hence there is a bias toward showing a relative decline.

We should note that the proper treatment of the biomass reporters and the shutdown or misreport dynamics does not impact the co-pollutants as much: in general, the introduction of a shift variable for the change in reporting and the treatment of shutdowns in a different way means that the coefficients reported in a regression along the lines of the third column of Table 1 for PM2.5, PM10, NOX, and SOX show a fall in absolute value but nowhere near as much as do the trend coefficients for GHGs, mostly because the problem of reporting co-pollutants after shutdown is not as severe (i.e., CEIDARS reporting does not include data when a facility has been closed) and the reporting shifts in PM in 2011 are less differential and so more easily captured by a fixed year effect. In any case, this is another instance in which more caution with the data would have been warranted.

Explaining Variation

In the body of the paper, we note that the common percentage estimation approach obscures heterogeneity that is at the heart of EJ concerns. To see the challenge, consider Table 3 in which we show four alternative scenarios, each achieving a 10 percent reduction in total regulated-sector emissions, but with very different distributions of emission changes across facilities. To illustrate HCM's identification of the policy effect by comparison to the non-regulated sector, the table includes a non-regulated facility, and there are two CT Facilities of which Facility 2 is the larger. There are three periods: pre-policy which is well before the policy goes into effect; the eve of the policy just as it goes into effect; and after the policy has had its effect.

For simplicity and without loss of generality, the example scenarios assume that the non-covered comparison facility does not change. We set the values for Facility 1 & 2 such that κ_1^p, the pre-policy trend, is equal to .05 in all cases (with that driven by Facility 1 experiencing a slight uptick in emissions on the eve of the policy). We then assume a global (or common) reduction of 10 percent but consider four scenarios. Scenario 1 achieves the global reduction target (or "cap") through balanced, across-the-board, common-percentage reductions at both CT Facilities. Scenarios 2 through 4 achieve the same global reduction target but through increasingly unbalanced percent reductions across facilities. Specifically, Scenario 2 keeps a constant difference in the physical quantity of releases with some reduction at each regulated facility, Scenario 3 leaves emissions from the larger facility unchanged with all reductions occurring at the other regulated facility, and Scenario 4 has an increase in emissions at the larger facility offset by large reductions at the other facility.

Table 3. Four Emissions Reductions Scenarios (version 1)

Four scenarios, each achieving a 10 percent reduction in total regulated-sector emissions but with different distributions of change across facilities

	Scenario 1			**Scenario 2**			**Scenario 3**			**Scenario 4**		
	Non CT	CT Facilities		Non CT	CT Facilities		Non CT	CT Facilities		Non CT	CT Facilities	
	Facility	Facility 1	Facility 2	Facility	Facility 1	Facility 2	Facility	Facility 1	Facility 2	Facility	Facility 1	Facility 2
pre-policy	100	90	200	100	90	200	100	90	200	100	90	200
eve of policy	100	100	200	100	100	200	100	100	200	100	100	200
after policy takes effect	100	90	180	100	85	185	100	70	200	100	60	210

Measures of change in relative disparity, overall emissions, and emissions gap

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Relative CT / Non-CT after policy	90%	90%	90%	90%
Change Average CT emissions after policy	-10%	-10%	-10%	-10%
Change in Emissions Gap after policy	-10	0	30	50

If you apply the average change to both Facility 1 & Facility 2

	Scenario 1			**Scenario 2**			**Scenario 3**			**Scenario 4**		
	Non CT	CT Facilities		Non CT	CT Facilities		Non CT	CT Facilities		Non CT	CT Facilities	
	Facility	Facility 1	Facility 2	Facility	Facility 1	Facility 2	Facility	Facility 1	Facility 2	Facility	Facility 1	Facility 2
Eve of policy	100	100	200	100	100	200	100	100	200	100	100	200
Estimated change after policy	100	90	180	100	90	180	100	90	180	100	90	180
Estimated change in Emissions Gap after policy		-10			-10			-10			-10	

We say this is version 1 because the actual pattern is a bit more complicated than we portray but Table 3 helps get across the basic intuition: if you apply a common percentage, each scenario yields the same estimated improvement in the EJ gap while the actual variation between the scenarios is quite large. For analysts interested in what may have actually happened to the EJ gap, a common percentage estimate could be quite misleading.

Table 4 offers a more complicated view that suggests even more potential issues with the HCM approach. Here, we simulate four cases and calculate the full post-policy trend ($\kappa_1^p + \kappa_2^p$) using a formula that makes use of the asinh values, which we then use to calculate the implied trend break. As before, in all cases, the total or global reduction is on the order of ten percentage points; note, however, that the greater the variation across facilities in response to cap and trade, the higher the HCM-style estimate of κ_2^p , mostly due to the way in which asinh changes are calculated.

Table 4. Four Emissions Scenarios (version 2)

Four scenarios, each achieving a 10 percent reduction in total regulated-sector emissions but with different distributions of change across facilities

	Scenario 1			**Scenario 2**			**Scenario 3**			**Scenario 4**		
	Non CT	CT Facilities		Non CT	CT Facilities		Non CT	CT Facilities		Non CT	CT Facilities	
	Facility	Facility 1	Facility 2	Facility	Facility 1	Facility 2	Facility	Facility 1	Facility 2	Facility	Facility 1	Facility 2
pre-policy	100	90	200	100	90	200	100	90	200	100	90	200
eve of policy	100	100	200	100	100	200	100	100	200	100	100	200
after policy takes effect	100	90	180	100	85	185	100	70	200	100	60	210
How the HCM equation 1 captures each scenario:*												
HCM's kappa1		0.05			0.05			0.05			0.05	
HCM's kappa2		-0.16			-0.17			-0.23			-0.28	
kappa1 + kappa2		-0.11			-0.12			-0.18			-0.23	
Measures of change in relative disparity, overall emissions, and emissions gap												
Relative CT / Non-CT after policy		90%			90%			90%			90%	
Change Total CT emissions after policy		-10%			-10%			-10%			-10%	
Change in Emissions Gap after policy		-10			0			30			50	
With an estimated 10% global decline, what goes into the HCM air plume model, by scenario												
	Non CT	CT Facilities		Non CT	CT Facilities		Non CT	CT Facilities		Non CT	CT Facilities	
	Facility	Facility 1	Facility 2	Facility	Facility 1	Facility 2	Facility	Facility 1	Facility 2	Facility	Facility 1	Facility 2
pre-policy	*100*	*93*	*193*	*100*	*92*	*196*	*100*	*88*	*205*	*100*	*85*	*212*
eve of policy	*100*	*98*	*204*	*100*	*97*	*206*	*100*	*93*	*216*	*100*	*90*	*223*
after policy takes effect	*100*	*88*	*183*	*100*	*86*	*183*	*100*	*77*	*181*	*100*	*71*	*177*
Estimated change in Emissions Gap after policy		*-11*			*-12*			*-20*			*-28*	
Effect of Policy on the Emissions Gap between CT Facilities 1 and 2												
Reality		***Narrowing***			***Static***			***Widening***			***Widening***	
HCM Finding		***Narrowing***			***Narrowing More***			***Narrowing Even More***			***Narrowing Even More***	

* We assume in all cases that there is a ten percent reduction in CT emissions relative to the non-CT sector. Note that HCM's kappa2 (and kappa1 + kappa2) is larger in magnitude the more unbalanced the adjustment across facilities. If that percentage effect is then applied to estimates, this will create a bias toward finding closure of the emissions gap when it may have widened.

In the bottom half of Table 4, we show the estimated declines from an HCM-style common percentage estimation. Note first that the initial estimated starting points for each of the C&T facilities is not the same as the "real" starting points; that is because the regression procedure means that each facility starting point includes the constant (which is the non-C&T starting point) and a fixed effect that is estimated over all three years. Now track through the estimated changes in the pollutant and note once again that HCM's model cannot distinguish between the balanced, across-the-board reduction of Scenario 1, which reduces the physical emissions gap between the facilities, and the lesser or greater “hot-spotting” of Facility 2 in Scenarios 2, 3, and 4; indeed, because the greater the variation across facilities in response to cap and trade, the higher the HCM-style estimate of κ_2^p , we get a bigger estimated reduction in the emissions gap even as the emissions gap grows across the scenarios.

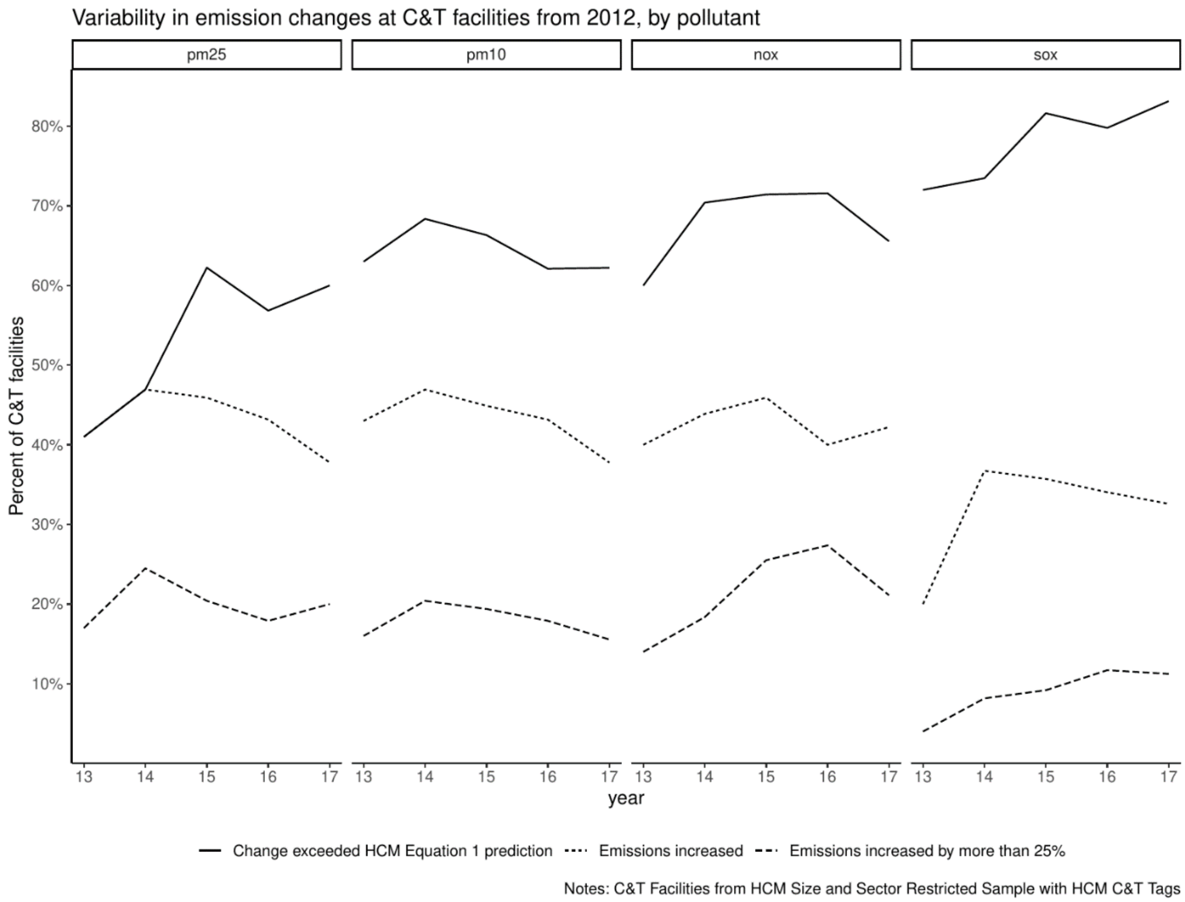
HCM include an appendix to show that one cannot work from facilities only to capture the actual EJ gap since facilities can affect both DAC and non-DAC neighborhoods; it is better to focus on the receptor side. We note below that we concur that this is a better approach. Still, it would seem that if Facility 2 has a plume that tends to affect more EJ communities than Facility 1, the common percentage modeling might be missing the impact on EJ communities. By design, cap-and-trade policy can generate environmental winners and losers, even if there is an average benefit to society. The HCM approach assumes that variation away.

It is also useful to stress again that an assumed common percentage effect necessarily predicts larger reductions in mass where there is more pollution. Since we know that there is generally a preexisting pattern of environmental disparity – a point both we and HCM make in our analyses (Hernandez-Cortes and Meng 2023; Pastor et al. 2022) – this prediction combined with an initial racial disparity in the distribution of pollution is biased towards “finding” (or rather estimating) larger reductions for EJ communities.

None of these considerations would be a big empirical issue if the variation in changes, and the sharp divergence of observed changes from the common-percentage model assumed and imposed by HCM was just a feature of our simple simulation and not a pattern in the data. To explore this, we focus on co-pollutants (since the GHG pattern is so problematic and the concerns are mostly about co-pollutants) and compare the predictions in HCM’s stage 1 model to predict the level of emissions (of any pollutant) for each facility and year, and then compare predicted changes for facilities from 2012, the eve of policy implementation, to year t, that is, $\widehat{Y_{\iota t}} - \widehat{Y_{\iota,2012}}$ to actual changes for facilities from 2012 to year t ($Y_{(i,t)}$– $Y_{(i,2012)}$). In generating these predicted values, we utilize the C&T tags that HCM use; as we have noted, these may be problematic but here we are comparing what they predict to what actually occurred.

First, we consider the variation across the cap-and-trade facilities that appear in the HCM regression, i.e., with the size and sector limitations, in the changes in emissions over time. Figure 6 shows three time series for each pollutant. We focus on the four co-pollutants, PM2.5, PM10, NOx, and SOx, rather than GHG because co-pollutants are relevant for the analysis of inter-facility variability and environmental justice, and because of the earlier indicated issues with the modeling of GHG when one does not account for the reporting shift that occurred in 2011.

Figure 6

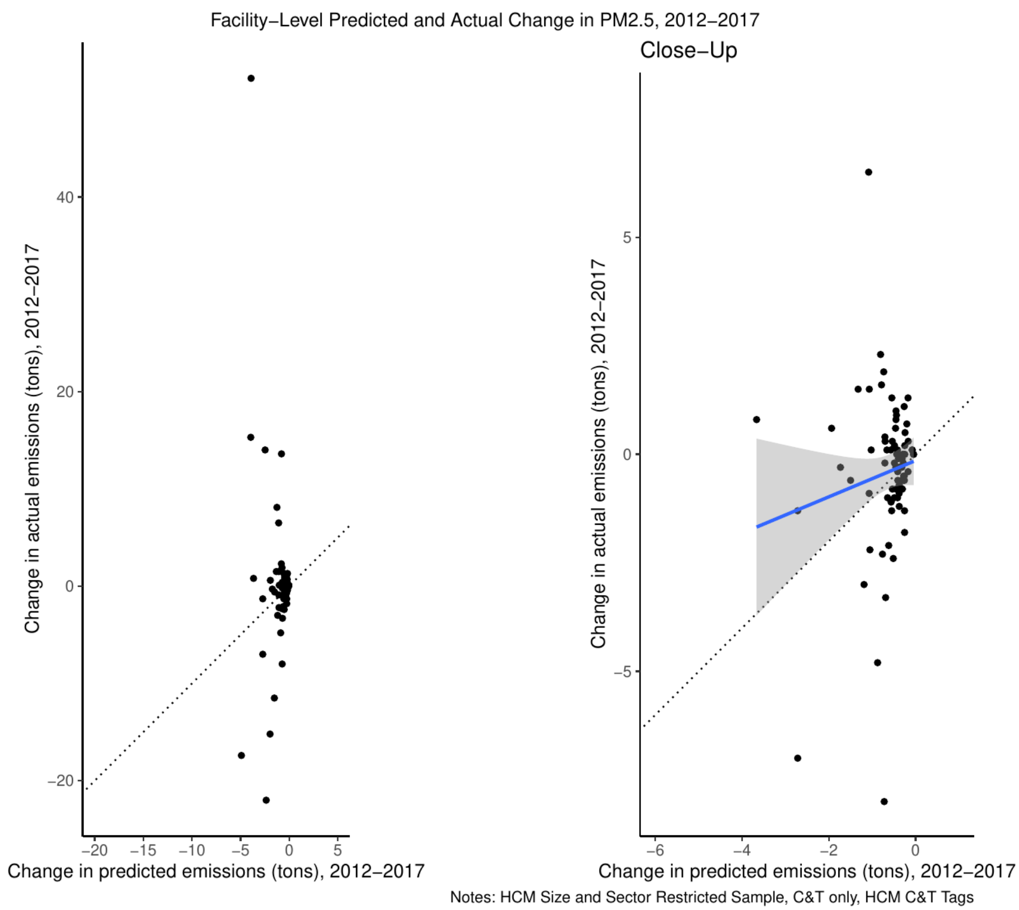


The first series is the share of cap-and-trade facilities whose actual change from 2012 to the indicated year exceeded the predicted change. Predictions based on a regression model of course will of course have error, but a good model will generate predictions that are dispersed both above and below the predicted value. That is we would expect roughly 50% of facilities to have actual changes above the predicted change (and 50% below). In fact, for all pollutants for almost all years, the percent of facilities that have actual changes above the predicted change is at least 60%, rising to 70% for NOx in 2014–2016 and around 80% for SOx for 2015–2017. The high share of facilities with actual changes that exceeds predicted changes suggests either outright bias in the estimation equation or, at best, a highly uneven distribution of values, which could, for example, reflect a small number of facilities with large emission declines (or even closures) offsetting wide incidence of increasing or stable emissions,

The second line in the time series plot reports the percent of facilities with actual changes greater than zero over the period considered. This share is between 40% and 50% for PM2.5, PM10, and NOx and is somewhat lower for SOx, for which it finishes the period at roughly 40%. Finally, the third time series reports the share of facilities with increases of at least 25% increase in emissions. In most years, more than 20% of facilities had experienced increases relative to 2012 of more than 25%. Thus, nearly half of facilities actually experience an increase in emissions and 20% of facilities experience substantial increases in emissions. These are non-trivial deviations from the common-percentage predicted reduction in emissions HDM finds, which indicates that the reality may more closely resemble scenarios in which changes in emissions vary sharply across facilities (scenarios 3 and 4 in Table 3) rather than one in which all facilities experience across the board common-percentage changes.

Next, we present in Figure 7 a scatterplot for PM2.5 in which the actual changes between 2012 and 2017 for each facility is plotted against the change predicted by the HCM common-percentage model for that facility. We start with PM2.5 here because it is the pattern in which the HCM modeling gets closest to the pattern and we wish to offer their best case. Although we focus on the change from 2012 to 2017, the results are similar for changes from 2012 to any year after the implementation of the policy, which is not surprising given the evidence on variation presented in Figure 6.

Figure 7



In the scatterplot, the 45° line indicates what would be the case if equation 1 perfectly predicted facility-level quantity changes. For example, a facility predicted to decrease emissions of PM2.5 by 2 tons that in fact decreased emissions of PM2.5 by 2 tons would appear on the 45° line. If this facility reduced PM2.5 emissions by only 1 ton, the point for the facility would appear above the 45° line, and if it reduced PM2.5 emissions by more than 1 ton, the point for the facility would appear below the 45° line. The left-hand panel shows the full scatterplot of actual changes versus predicted changes for the 90 cap-and-trade facilities in the HCM regression sample that had data for 2012 and 2017; because the spread of actual changes is very high, panel (b) shows a close-up of observations closest to the origin (deleting 9 outliers

Three points are clear. First, the spread of actual changes is very large relative to the spread of predicted changes from the common-percentage model. Second, the points are generally above the 45° line, supporting the implication of substantial bias in the HCM model that was suggested by the asymmetry between over-prediction and under-prediction in Figure 6. Third the actual change is not closely or even broadly predicted by the predicted change even in direction, let alone magnitude. The bivariate regression lines from a regression of the actual change on the predicted change have a slope that diverges from the 45° line that would be generated by an accurate but merely imprecise model.

The EJ Gap results that emerge from stage 2 of the HMC analysis depend crucially on the facility-level emissions changes that go from stage 1 into the stage 2 model, and HCM have entered values from the 45° line into stage 2. These values seem to bear little relation to the actual performance of cap-and-trade facilities after cap-and-trade went into effect. It is crucial to emphasize that this is not merely a case of statistical variation around a fundamentally accurate trend. The variation is wide and we suggest that it is the variation that is crucial to capture not the average effect.

Below, we show the contrast in the actual values for the change in 2012-2017 vs. the predicted values for PM2.5 for 2012-2017 that would emerge using the HCM C&T tags, the HCM regression method, and the HCM subsample for GHG and the three other co-pollutants in a series of four figures. For GHG, the values for change has fewer outliers; for the other co-pollutants, there are outliers and so we take the approach as in the consideration of PM2.5 and also offer a close-up view of the data.

Figure 8. Predicted and Actual Change, Total GHG

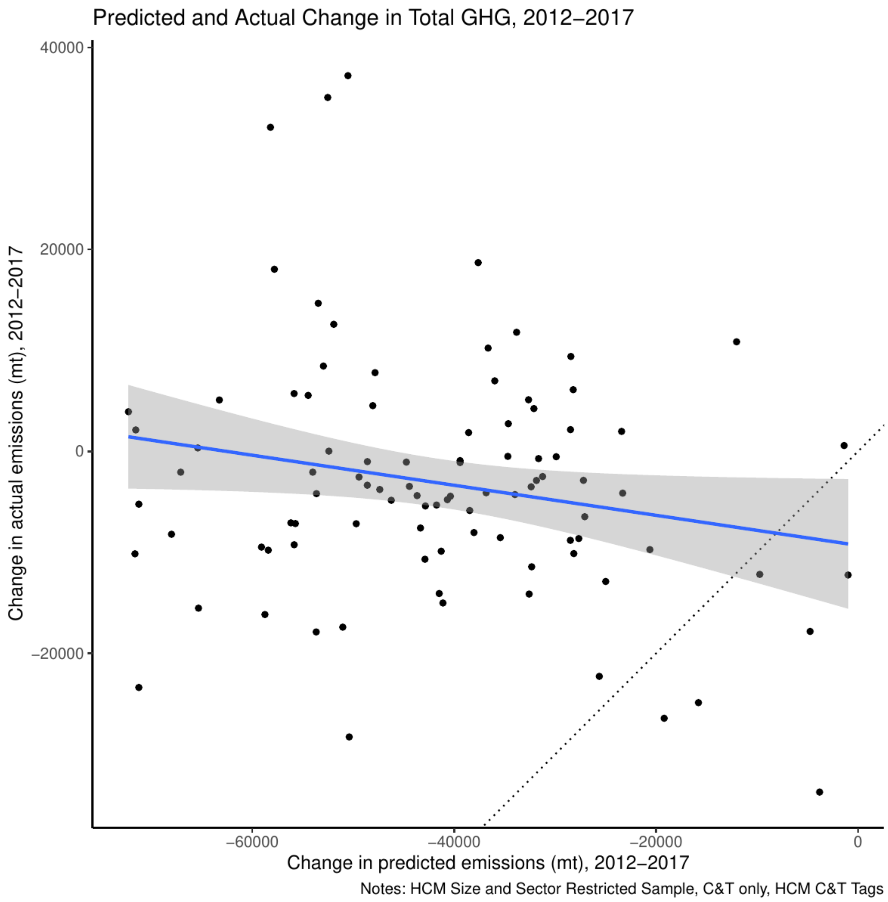


Figure 9. Predicted and Actual Change, PM10

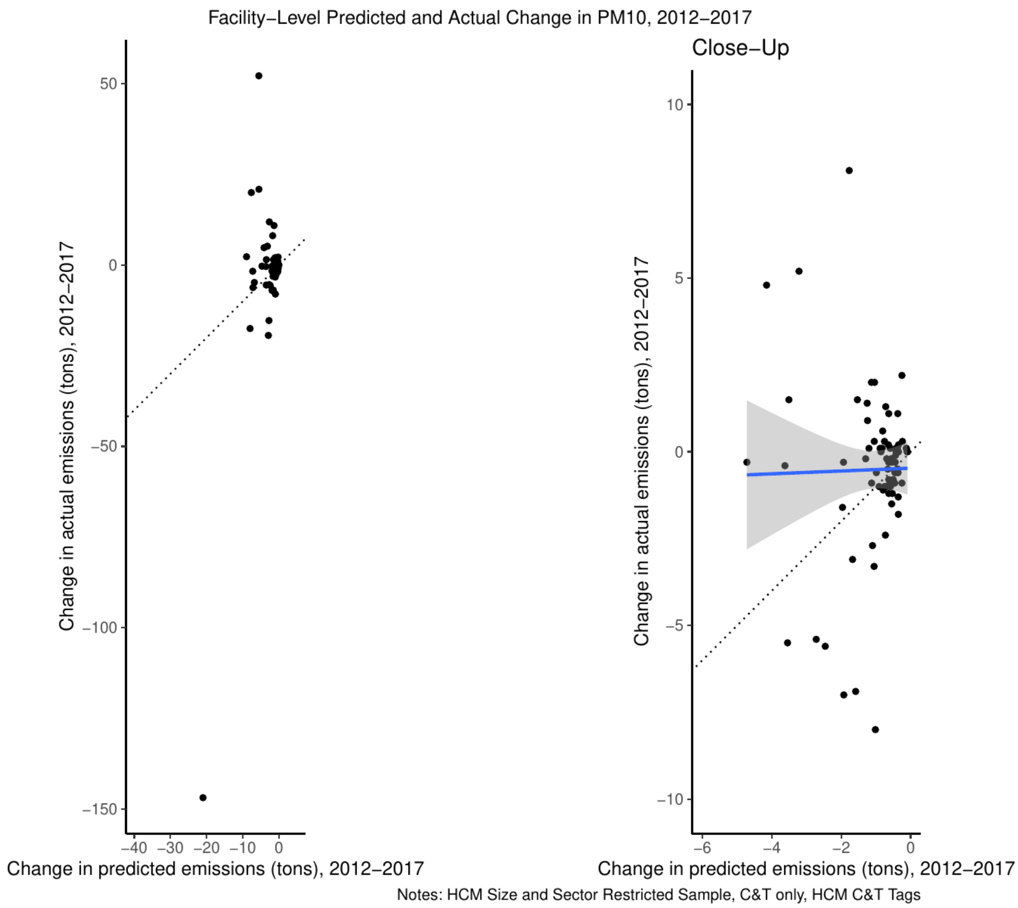


Figure 10. Predicted and Actual Change, NOX

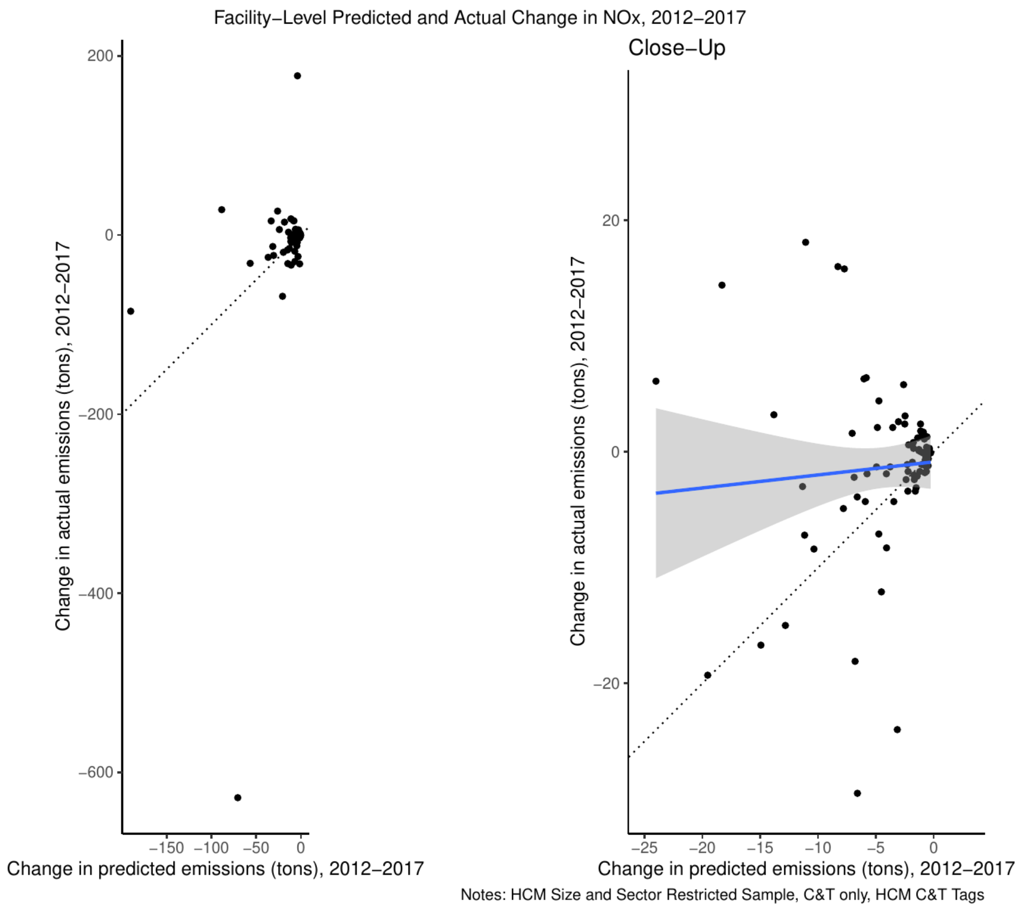
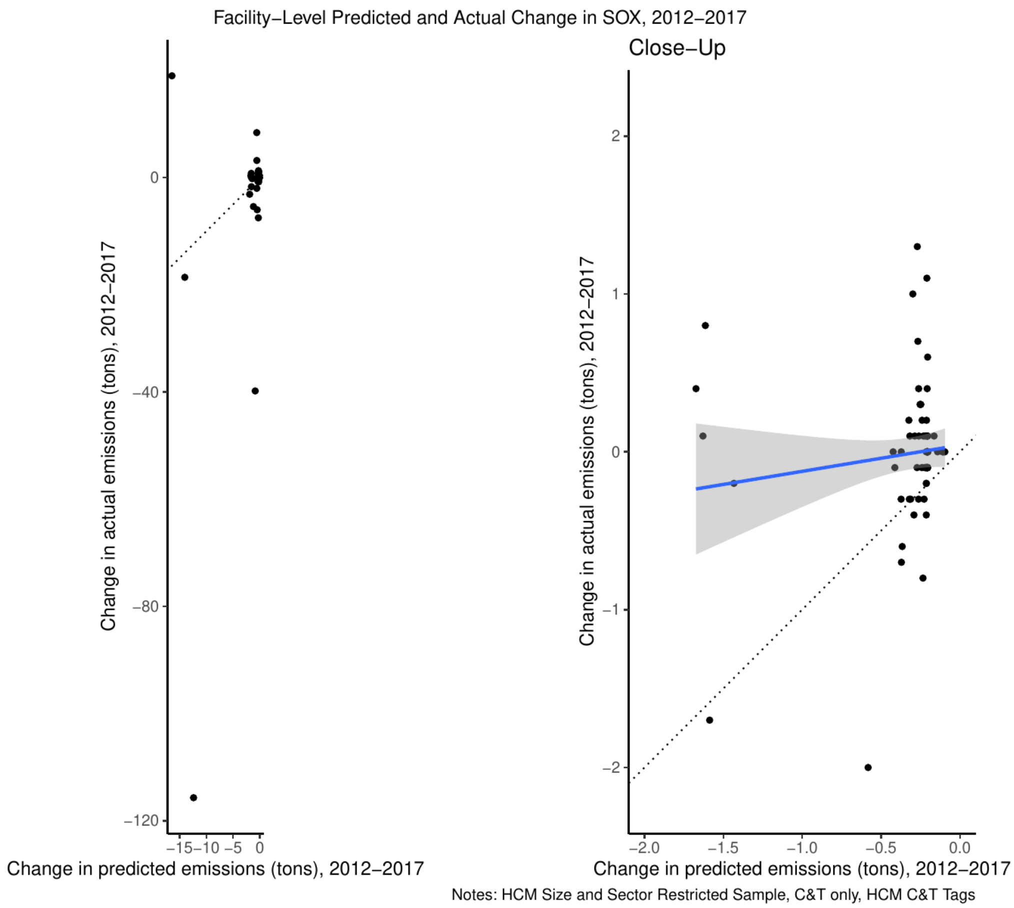


Figure 11. Predicted and Actual Change, SOX



One thing that is quickly noticeable – and might have been surmised from the extensive discussion above – is just how poor the fit is for GHGs. We can see, however, that PM2.5 comes closest to being considered reasonably estimated, and the fit seems quite unimpressive for the other co-pollutants depicted in this appendix.

Another Approach to Understanding Relative Performance

While HCM rightly object to a before-and-after approach that does not include a control group (Hernandez-Cortes and Meng 2022), one can conduct an event-study analysis of the difference since 2012 between the change in mean inverse hyperbolic sine of cap-and-trade facility emissions and the change in mean inverse hyperbolic sine of non-covered facility emissions. An event study analysis showing facility changes relative to 2012, the eve of policy, requires the availability of data for 2012 and one other year to contribute to identification, a minimal inclusion criterion to identify policy-relevant coefficients from within-facility changes. We focus on co-pollutants because of serious issues with the estimation of the GHG patterns detailed above.

Figure 12. Event Study Analysis for the HCM Restricted Sample

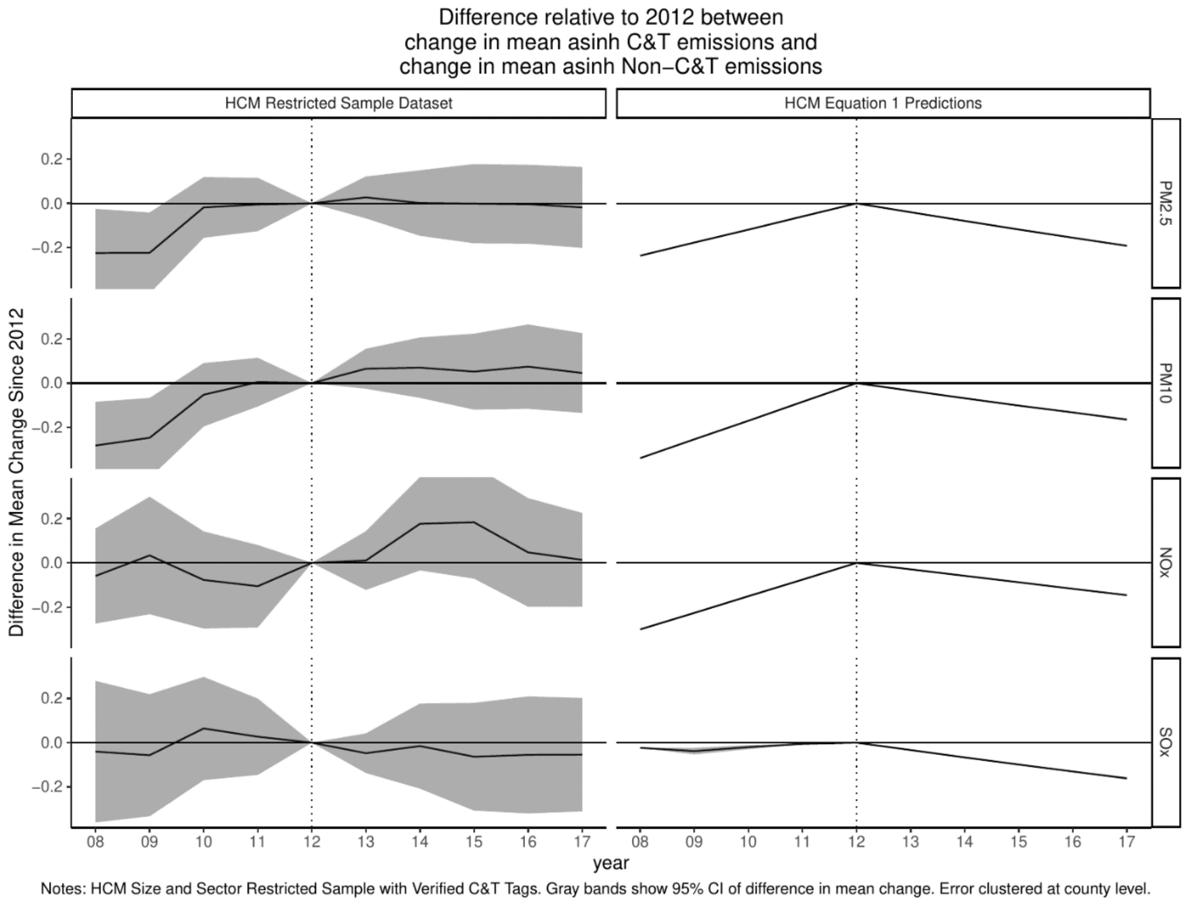
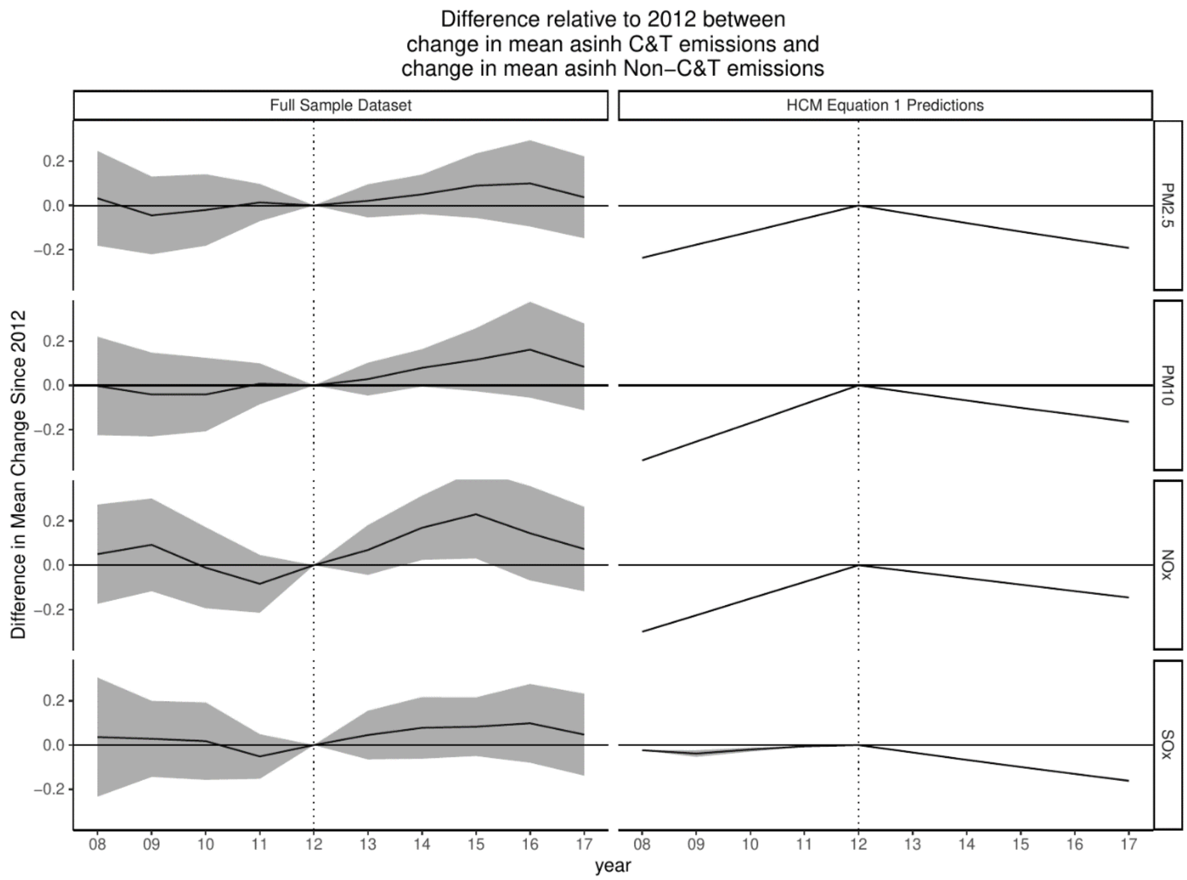


Figure 12 focuses on the HCM sample that is restricted by size and sector to approximately 5% of GHG emissions, but we correct the Cap-and-Trade designations, as we do in Column 2 of our regression re-analysis in Table 1. The right-hand panel shows what the HCM regression would have predicted for each co-pollutant; naturally there is no variation but simply a steady trend (recall that by focusing on the pattern relative to the control group, we are stripping away the year effects that showed up, for example, in Figure 4 on GHGs). The left-hand panel shows the pattern from the actual data with the gray bands present a 95% confidence interval around the point estimate. In no year does the event history analysis find a statistically significant effect of cap-and-trade, either positive or negative, in terms of the gap in growth between C&T and non-covered facilities in the years after the policy goes into effect. This corroborates column 2 of our regression re-analysis and contradicts the HCM reported finding for equation 1.

We can also see that event study analysis of the years before 2012 suggests a pre-policy upward trend difference for PM10 and PM2.5. While this might be heartening for the HCM regression, it is important to note that there was a change in reporting standards specific to PM10 and PM2.5 starting in the year 2011 that CARB warns “caused an increase in PM, PM10, and PM2.5 reported emissions” (See the Caveats documents at: https://ww3.arb.ca.gov/carbapps/pollution-map/). While this might be captured by shared year effects, an examination of the data in the HCM subsample suggests that the impact may have differed by whether a facility was cap-and-trade or not, making the upward relative pre-policy trend less reliable than the relatively flat trends shown for NOX and SOX (Pastor et al. 2022:38).

We can also use the event study approach to examine the impact of HCM's subsample which eliminates facilities accounting for 95% of emissions in search of best comparison groups. One way to do that, akin to the fifth column of Table 1, is to compare the HCM estimate from their subsample to the event pattern for the full universe of cap-and-trade facilities and non-covered facilities.

Figure 13. Event Study for the Full Dataset



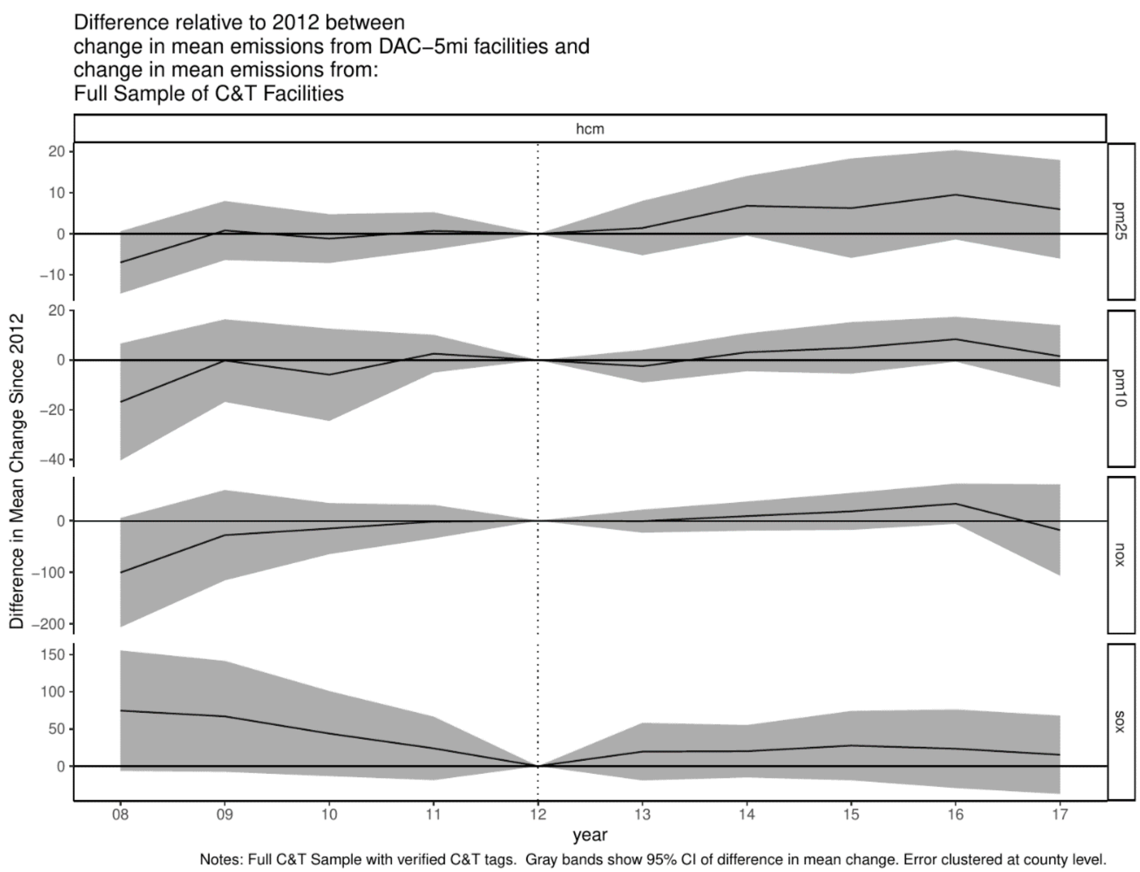
Notes: Full Sample with Verified C&T Tags. Gray bands show 95% CI of difference in mean change. Error clustered at county level.

As in the fifth regression in Table 1, when we include the full set of polluting facilities in the event study, we find less indication of pre-policy trend differences. Furthermore, the post-policy trend difference indicates that emissions increased in the cap-and-trade sector relative to the non-covered sector, although the differences are for the most part not statistically significant.

A final exercise involves looking at whether there is a distinct pattern for facilities that tend to disproportionately impact DACs. HCM are right to insist in an appendix that their two-stage approach – first estimate the impact of cap-and-trade, then run those estimates through an air model – may have advantages over a more common approach of ascribing demographic information to the facility (either through proximity analysis or through aggregating from the neighborhoods affected by the air plume). However, a careful reading of that appendix suggests that a facility-based approach points in a different direction than a neighborhood approach in the circumstance in which the EJ gap is negative – whiter and wealthier communities are more polluted than low-income communities of color. In that case, a slight relative increase from facilities affecting the latter communities can be overwhelmed by a sizeable increase affecting all communities – in which case the EJ gap becomes even more biased against advantaged communities.

That is a possibility, of course, and one should be open to not finding a pre-existing pattern of uneven exposure related to race and income. Still, both HCM and we (and nearly all observers) agree that the DACs face more initial exposure from facilities. While it remains better to take a two-step approach, with this data, a source and receptor analysis should generally point in the same direction. In Figure 14, we compare the pattern for the full sample for facilities that have a DAC within 5 miles as compared to those that do not for the four co-pollutants. We did other comparisons for the HCM size- and sector-constrained samples and for different buffer distances. The bottom line is that there is no convincing evidence of relative improvement for the DAC-impacting facilities.

Figure 14. Event Study Comparing DAC and non-DAC Facilities



A Brief Recap

In the main body of this paper, we noted a series of analytic and data issues regarding whether a common percentage approach addresses EJ concerns, whether the facilities that HCM use in their analysis are correctly coded as to cap-and-trade status, whether the imbalance in the sample – particularly the obvious bunching of entries into the database and the fact that the majority of facilities in their regression subsample lack observations before and after policy implementation – can lead to shifting results, and whether one can correct for the imbalance by linking in the original CEIDARS data. We also suggested that it is not clear that the HCM results are robust across the whole sample.

This appendix highlights how the coefficient estimates and the predicted values from the HCM specification for GHG might be distorted by a reporting shift that CARB warns researchers to consider. We also discussed how facility closures unrelated to the cap-and-trade policy can impact the results and have a larger effect on estimates of the trend break for the C&T facilities that then potential overstates the post-implementation impact of the policy. We also explored how the fit between the HCM-predicted changes in pollutants between 2012-2017 do not seem to dovetail well with the actual pattern.

Finally, we showed how the HCM-predicted estimates vary from the results obtained from such an event study approach that does include a control group and uses the verified C&T tags from CARB/OEHHA, showing divergences that are particularly large for the full sample. We concur that air modeling is a potential step forward for the analysis of EJ issues but this second stage depends critically on the estimates that emerge from a first stage. The database that is used to estimate the changes in predicted pollutants is not large and it is worth the time to explore the sensitivity of results to better specifying and resolving these data issues, as well as considering other analytical approaches.