The Impact of Aid on Deforestation*

Yota Eilers†, Jochen Kluve‡, Johannes Schielein§ Melvin H.L. Wong ¶

August 31, 2022

We show that development finance of forest protection areas in Latin America effectly reduces the speed of deforestation.

PAs are key to fighting global climate change and the biodiversity crisis (IPCC, IPBES) and it is planned to increase PA coverage from currently about 17% to 30% of land surface area.

Little robust evidence on conservation effectiveness, especially at the portfolio level.

Problem: Monitoring data might suggest that we are not successfully reducing forest cover loss (example figures right, project start at red line). Monitoring data are not based on a robust counterfactual scenario (what would have happened without our effort?) => Monitoring might therefore underestimate conservation impacts, and lead to wrong conclusions about project effectiveness Implications: Low trust in forest cover protection measures (e.g. REDD+) => fewer investments in forest conservation finance => more loss + emissions Main research question Was forest cover loss effectively reduced in KfW supported protected areas in Latin America between 2004 and 2020?

1 Data

KfW is one of the most important conservation financiers globally Ongoing portfolio (2021): 2.6 Bio. Euro, 602 Supported Protected Areas (PAs) in 66 countries

2 Empirical Strategy

The major challenge to identify conservation effects of protected areas is to account for selection bias. The protection of forest areas are not assigned randomly but policymakers make conscious decisions to declare protected areas. A major selection bias is the remoteness of an area. Forest areas that are distant to settlements have a lower probability of deforestation since the population to harvest timber is absent. Likewise, machines, such as harvesters, have a greater challenge to reach rugged forest areas. In contrast, areas with soil highly suitable for agriculture may be subject to increased deforestation pressures.

To overcome this selction bias we use Coarsened Exact Matching (CEM) to create statistical twins across our units of observations (Iacus et al., 2012; Blackwell et al., 2009). CEM is a matching method that relies on pruning your data so that each data point has a relevant counterpart. As an example, matching treatment and control units by travel time to the nearest settlement with CEM involves: First, defining range bins of the variable (e.g. 10, 30, 60, 90 minutes). Second, keeping only those observations in bins with treatment and control units. That is, if units in the bin of 90 minutes travel time and above consists only of control units or only of treatment units, the observations would be discarded from the analysis. Third, if there happen to be an unequal number of treatment and control units in a bin weights are calculated. Finally, the pruned data may be used for further analysis, such as weighted regressions.

CEM is especially suitable in our setting. When applying CEM one has to consider the trade-off between keeping units of observations and increasing similarity of treatment and control units. The narrower the bins defined, the more comparable are treatment and control units. However, narrow bins may be wasteful since more units are likely to be discarded from the dataset. Similarly, wide bins make it more likely that treatment and control units fall in the same bin, thus keeping more observations. Still, treatment and control units may remain systematically different from each other after matching with wide bins. The ideal CEM scenario to prune the data via narrow bins for a large number of treatment and control units. The availability

*The findings, interpretations, and conclusions expressed in this article are entirely those of the authors. They do not necessarily represent the views of KfW Development Bank and its affiliated organizations.

†yota.eilers@kfw.de
‡jochen.kluve@kfw.de
§johannes.schielein@kfw.de
¶melvin.wong@kfw.de

of high resolution geospatial data allows us to define narrow variable bins while keeping a large number of observations as well as achieve a high comparability between treatment and control units.

CEM is applies applied once at the beginning of the intervention year. That is, if the disbursements to protected areas started in 2015, the CEM will consider the matching variables only at year 2015. Table 1 lists the variables used to match control cells to treatment cells along with their justification for matching. The majority of the matching variables are time-invariant with the exception of the forest cover area for the previous 3 years. This variable accounts for different levels of forest cover before the intervention and is crucial to account for systematic bias. Otherwise, it would be possible to compare the deforestation rates in forest areas with urban areas.

Table 1: Overview of matching variables

Covariate	Data source	Data description	Rationale
Travel time to next city (population >5,000)	Weiss et al. (2018)	Accessibility is the ease with which larger cities can be reached from a certain location. This resource represents the travel time to major cities in the year 2015. Encoded as minutes, representing the time needed to reach that particular cell from nearby city of target population range. We use the travel time to cities with a population of 5,000 to 110 million.	Proxy for infrastructure accessibility (main determinant of forest cover loss/deforestation)
Forest cover	Hansen et al. (2013)	"Tree cover in the year 2000, defined as canopy closure for all vegetation taller than 5m in height. Encoded as a percentage per output grid cell, in the range 0âĂŞ100."	Compare cells with similar level of forest cover and exclude non-forest areas (e.g. cities)
Forest cover loss (aggregated over three years prior to project start)	Hansen et al. (2013)	"Forest loss during the period 2000âĂŞ2020, defined as a stand-replacement disturbance, or a change from a forest to non-forest state. Encoded as either 0 (no loss) or else a value in the range 1âĂŞ20, representing loss detected primarily in the year 2001âĂŞ2020, respectively."	Captures pre-treatment deforestation dynamics
Terrain ruggedness	Farr et al. (2007); Riley et al. (1999)	This index is calculated using the same data as the elevation variable. The elevation difference between the center pixel and its eight immediate pixels are squared and then averaged and its square root is taken to get the TRI value. This function allows to efficiently calculate terrain ruggedness index (tri) statistics for polygons. For each polygon, the desired statistic/s (mean, median or sd) is/are returned.	Agricultural mechanization suitability
Elevation	Farr et al. (2007)	The layer represents the 30m global terrestrial digital elevation model from the NASA Shuttle Radar Topographic Mission (SRTM), available for download as 5 degree x 5 degree tiles. It is encoded as meter, representing the elevation at the particular grid cell.	Proxy for agricultural climate suitability
Soil clay content	Hengl et al. (2017)	Proportion of clay particles < 0.002 mm in the fine earth fraction (g/100g)	Agricultural soil suitability
Country		Corresponding country of the gridcell.	Political and regulatory framework

After the matching approach we construct a panel dataset to estimate the aid effectiveness to reduce deforestation in protected areas:

$$loss_{c,t} = \beta_1 T_{c,t} + \delta_c + \tau_t + \epsilon_{c,t}, \tag{1}$$

where $loss_{c,t}$ is the change in forest cover in cell c, and year t. $T_{c,t}$ a dummy variable being 1 since the year of KfW disbursements, and 0 otherwise. All regressions include cell fixed effects and year fixed effects to account for time-invariant heterogeneity and time-variant covariates that are common across all units of observations, respectively. Moreover, the regressions are clustered on the level of protected areas, where the control group in each country forms a single cluster.

Our estimates are based on multiple panel estimates, which we refer to as "matching frames." A matching frame is a common longitudinal data with the only difference that it includes observations only from the treatment and respective control group. Matching frames are constructed by selecting the start year of the intervention and perform the CEM for the treatment cells of that start year against all potential control cells. After that, all unmatched units of observations are discarded and time-variant covariates are merged with the matched units of observations. This process is repeated for the next year. Matching frames are necessary, as the treatment cells vary with each year and are, hence, matched to different set of control cells. In the following, the matching performance is illustrated for the year 2015 while the regression results are shown for each matching frame with valid parallel trends for the cell's forest cover area in hectres.

Figure 1: Overview of matching variable balancing

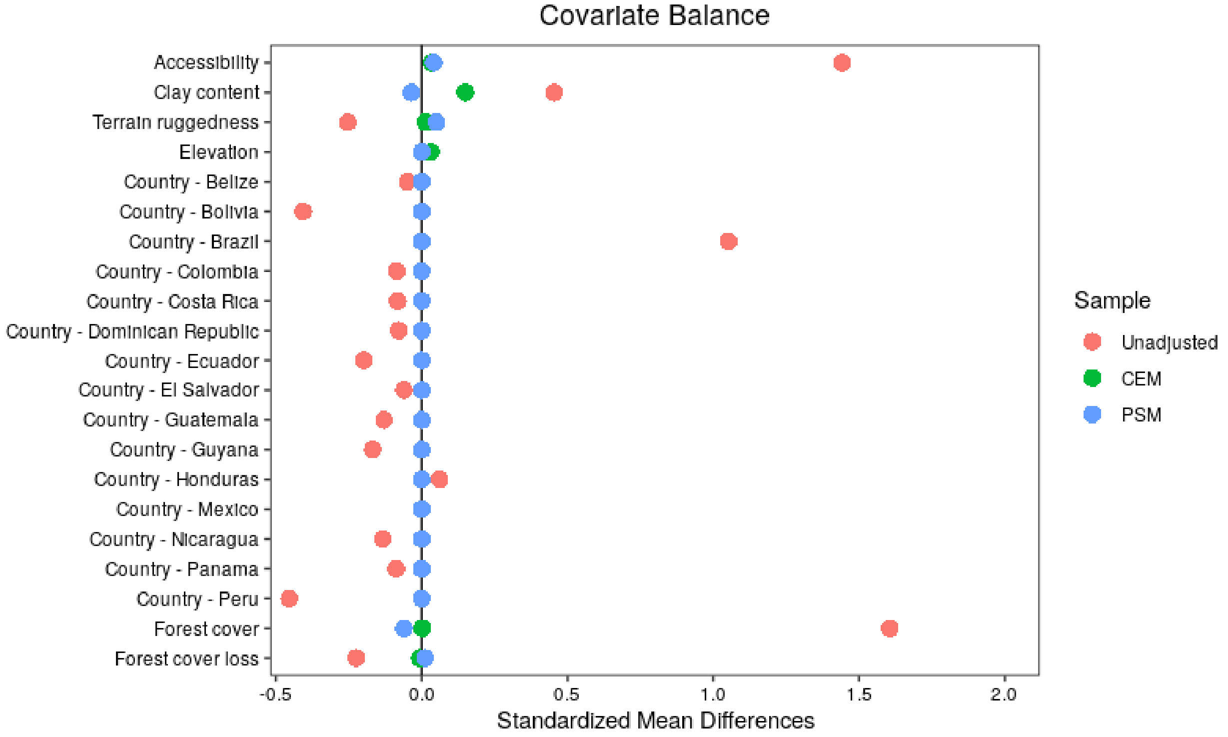
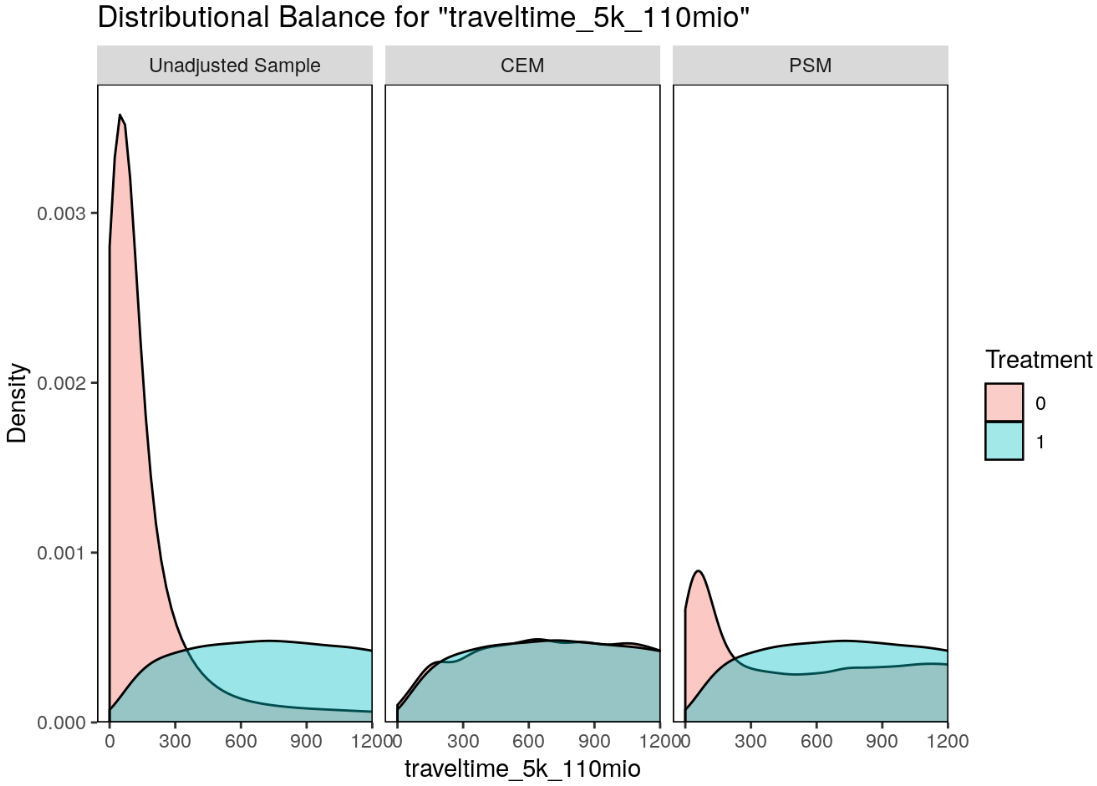


Figure 1 shows that matching establishes a balanced dataset as the mean differences between treatment and control group are significantly reduced and partially eliminated. Treatment and control group become especially comparable after matching with regards to travel distance to the nearest settlement and forest cover. As a further illustration, comparing the kernel densities of unmacthed and matched cells in figure 2 shows that distribution travel distances of treatment and control group become near identical for CEM matching and similar for PSM matching.

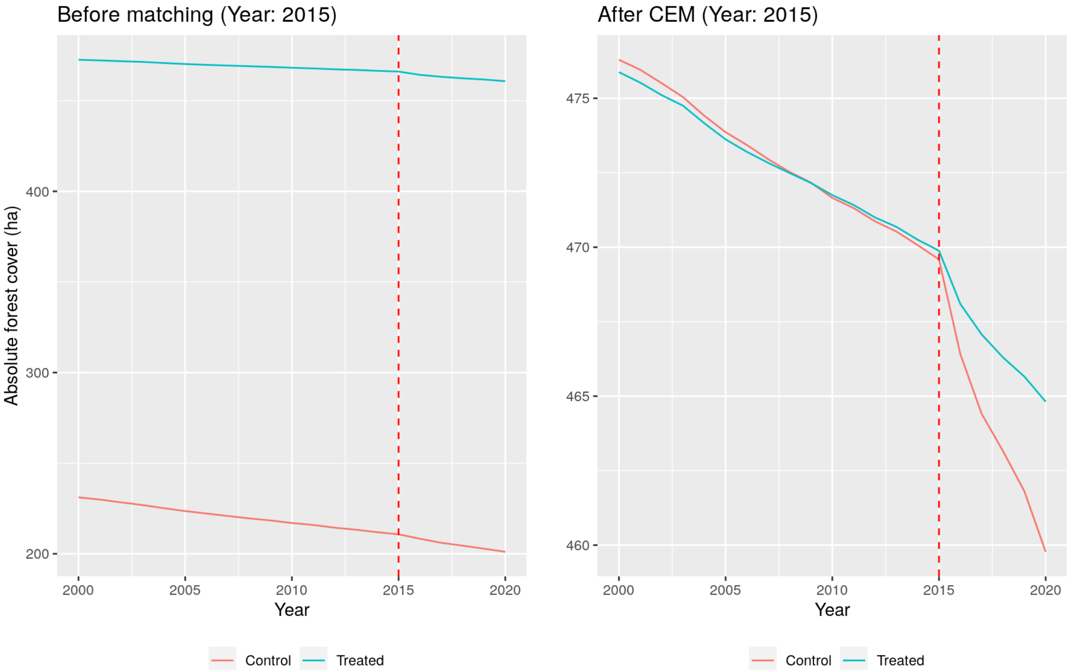
Figure 2: Distribution of travel time for treatment and control group before and after matching



3 Main results

Matching treatment and control groups

Figure 3: Parallel trends for the 2015 matching frame



Text about table

Table 2: Dependent variable: Forest cover loss

	2004	2005	2006	2007	2008	2009	2011	2012	2015
KfW support	-1.959***	-0.284	-0.198***	-0.462**	-0.122	0.703 +	-0.217	0.009	-0.767***
	(0.205)	(0.221)	(0.037)	(0.142)	(0.422)	(0.303)	(0.599)	(0.068)	(0.124)
Num.Obs.	4656920	5475260	3013860	12937100	5202440	286000	1379020	6576260	19885180
R2 Adj.	0.081	0.082	0.193	0.054	0.142	0.147	0.123	0.162	0.058
FE: .assetid	X	X	X	X	X	X	X	X	X
FE: year	X	X	X	X	X	X	X	X	X
Total num. cells	232846	273763	150693	646855	260122	14300	68951	328813	994259
–treated	300	933	11398	21862	9034	518	3007	11398	122433
–control	232546	272830	139295	624993	251088	13782	65944	317415	871826

Standard errors clustered by zones of forest protected areas (WDPA).
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

4 Conclusion

References

Blackwell, M., Iacus, S., King, G., and Porro, G. (2009). cem: Coarsened exact matching in stata. *The Stata Journal*, 9(4):524–546.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., et al. (2007). The shuttle radar topography mission. *Reviews of geophysics*, 45(2).

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., et al. (2013). High-resolution global maps of 21st-century forest cover change. *science*, 342(6160):850–853.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., et al. (2017). Soilgrids250m: Global gridded soil information based on machine learning. *PLoS one*, 12(2):e0169748.

Iacus, S. M., King, G., and Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 20(1):1–24.

Riley, S. J., DeGloria, S. D., and Elliot, R. (1999). Index that quantifies topographic heterogeneity. *intermountain Journal of sciences*, 5(1-4):23–27.

Weiss, D. J., Nelson, A., Gibson, H., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., et al. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553(7688):333–336.