Artisanal Mining Triggers Deforestation in the Democratic Republic of Congo

Supplementary Information, 15/12/2023

A Alternative data

The data used in our main analysis is based on Masolele et al. (n.d.). Although it relies on GFC data for the detection of deforestation events, it markedly differs from GFC in what it displays. In GFC, forest loss is essentially defined as tree cover loss, which can be a consequence of either deforestation or degradation (Pendrill et al. 2022). In contrast, the data from our main analysis focuses on post-forest land use and therefore on persistent land-cover change, which excludes many forest loss events in GFC.

To illustrate this difference, we reran the analysis with GFC data and compared it with results estimated with the Tropical Moist Forest (TMF) data set developed by (Vancutsem et al. 2021). TMF has the same resolution as GFC but does not work with a canopy threshold to define forest loss. Instead, TMF tracks forest changes since 1986 and classifies them as undisturbed, degraded or deforested forest pixels, depending on the severity and duration of canopy cover reduction. To be classified as deforested, disturbance within a forest pixel must exceed 2.5 years.

As the results in Figure 1 show, the deforestation estimates using TMF are similar to those from our main analysis, as they both detect persistent forest conversion, while TMF degradation and GFC forest loss are substantially higher. Our findings remain robust when alternative datasets are used, though degradation events captured by GFC are omitted.

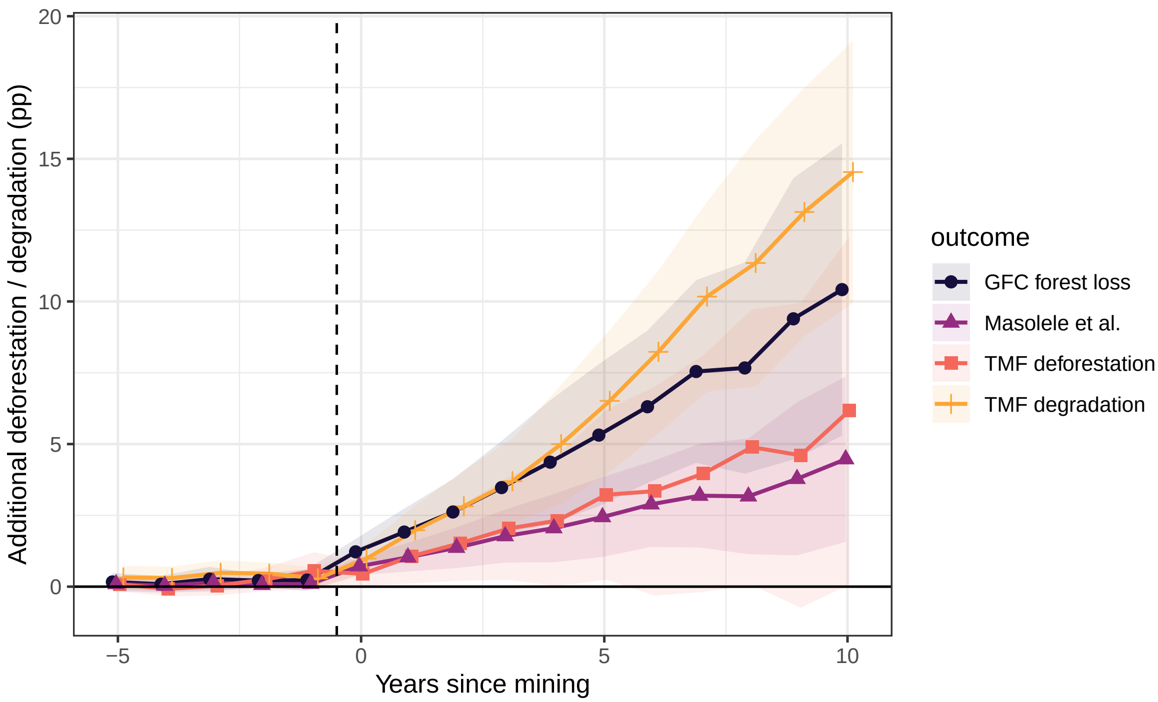

Figure 1: Additional deforestation or degradation effects by datasets, estimated following Callaway & Sant'Anna (2021).

B Alternative estimator

In addition to Callaway and Sant'Anna (2021) (CSA)'s group-time specific estimator, we also used the imputation (or two stage) estimators proposed by Gardner (2022) (GAR). In contrast to CSA, it follows a two-stage procedure. First, counterfactual trends that would occur in the absence of treatment are imputed by using the not-yet-treated units only:

$$Y_{it} = \mu_i + \lambda_t + u_{it} \tag{1}$$

Once Y is imputed, the ATTs can be estimated using a second stage regression of different treatment dummies on the imputed outcome.

Lead coefficients for the years before treatment can be included into the second-stage regression to test the plausibility of the parallel trends assumption. As Roth et al. (2022) notes, the GAR estimator relies on a more stringent parallel pre-trends assumption than that of CAS to obtain unbiased estimates, since it requires parallel pre-trends to hold for all included pre-treatment periods to impute the counterfactual. In contrast, CSA's group-time weighted average estimator only relies on parallel trends in the period before treatment begins for a valid counterfactual, although it seems unlikely that parallel trends would have been observed in the absence of treatment in such a case. Therefore, Borusyak, Jaravel, and Spiess (2022) suggest limiting the number of lead coefficients that are used to impute the first stage. In the absence of an established procedure by which to determine the number of leads, five pre-treatment periods were used throughout the analysis.

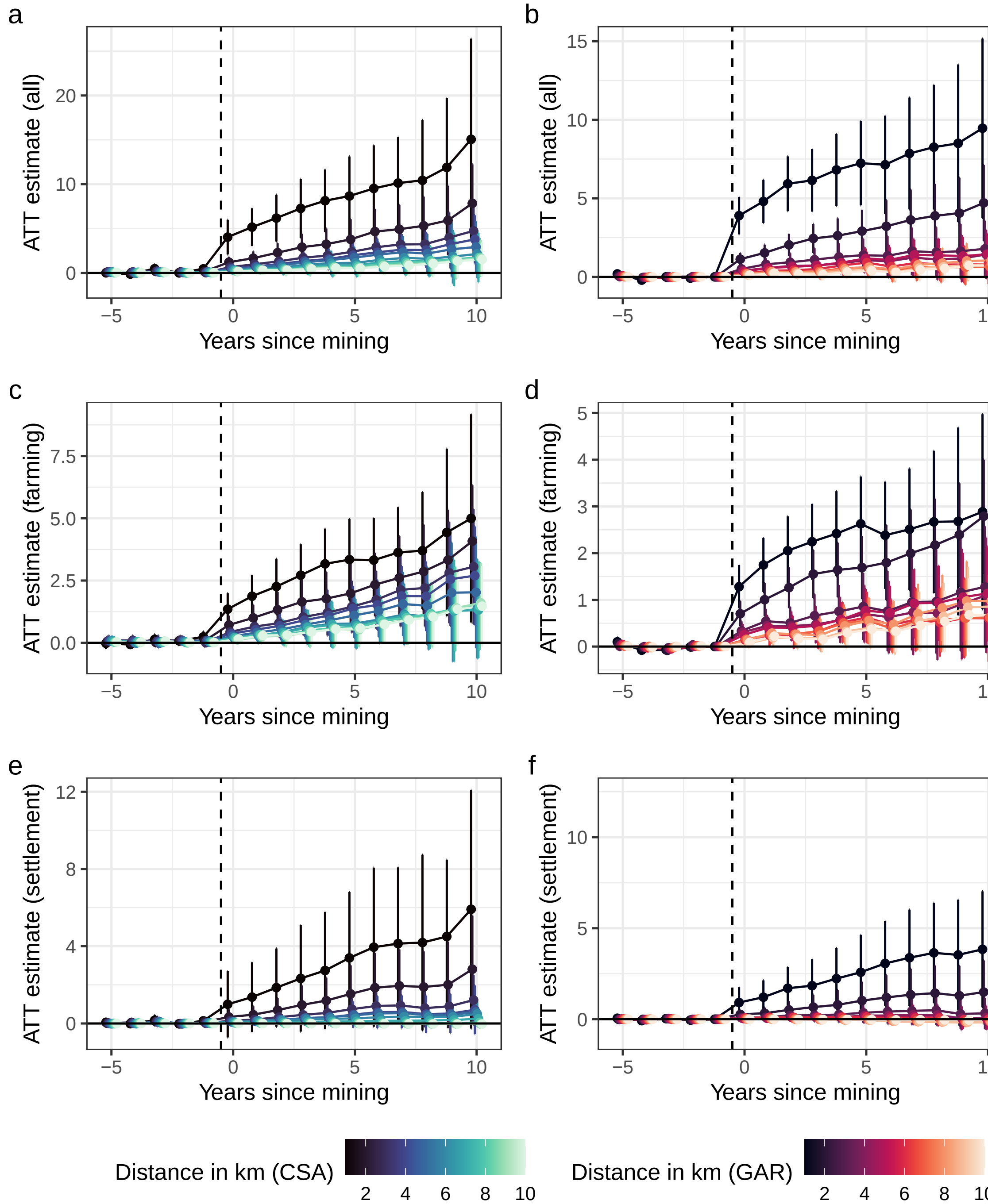

Figure 2: Event-study plots showing the ATT estimated as additional percentage points of forest lost within rings of increasing distance from the mining sites, over time. Plots (a), (c) and (e) are estimated following Callaway & Sant'Anna (2021), while (b), (d) and (e) employed the Gardner (2022) two stage procedure. The colours refer to the increasing distance from the mines. Confidence intervals are displayed at 95% level.

Figure 2 compares the CSA and GAR estimates in relation to time since the first mining incidence and distance from the mine. It shows that estimates run using GAR are more conservative across all specifications, although they still clearly reveal the effects of mining on surrounding forests.

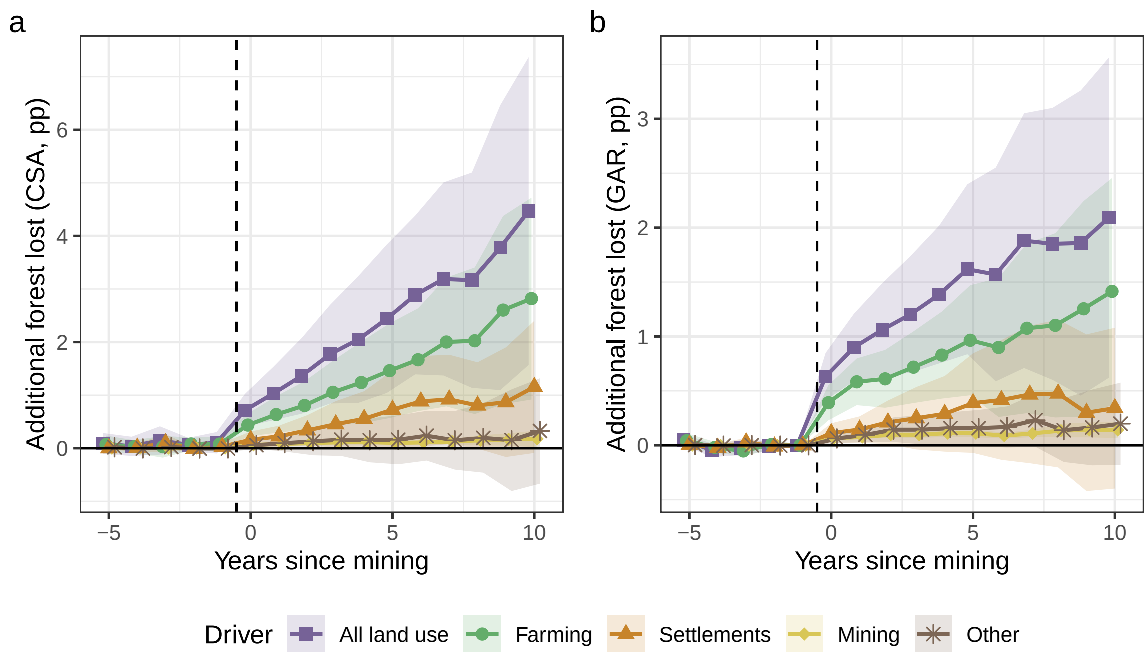

Figure 3: Estimates of impact on cumulative deforestation since 2000 as a share of forest cover in 2000 within a 5-km radius around mines for different periods following treatment, showing various land use following deforestation. For every additional kilometre from the mine, a new ring is estimated using (a) the estimator used by Callaway & Sant'Anna (2021), and (b) the estimator proposed in Gardner (2022).

Figure 3 compares the estimates for the CSA estimator, shown in panel A, and the GAR estimator, panel B. As with the spatio-temporal estimation, results from GAR are more conservative than from CSA, but still convey a similar message: mining leads to substantial forest loss within a 5-km buffer around sites, and farming activities play an important role in explaining the forest loss.

C Cluster threshold variations

	500m		1000m		1500m		3000m	
period	ATT	SE	ATT	SE	ATT	SE	ATT	SE
-5	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
-4	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.1
-3	0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.0
-2	0.1	0.0	0.1	0.0	0.1	0.0	0.1*	0.0
-1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1
0	0.8*	0.1	0.7*	0.1	0.5*	0.1	0.6*	0.1
1	1.2*	0.2	1.0*	0.2	0.8*	0.2	0.9*	0.2
2	1.5*	0.2	1.4*	0.3	1.1*	0.2	1.2*	0.2
3	2.0*	0.3	1.8*	0.3	1.5*	0.4	1.7*	0.3
4	2.5*	0.4	2.1*	0.4	1.7*	0.4	2.0*	0.4
5	2.9*	0.4	2.4*	0.5	1.9*	0.5	2.4*	0.4
6	3.5*	0.6	2.9*	0.5	2.3*	0.5	2.8*	0.5
7	3.9*	0.6	3.2*	0.7	2.5*	0.6	3.2*	0.6
8	4.1*	0.6	3.2*	0.7	2.3*	0.8	3.1*	0.7
9	4.7*	0.9	3.8*	1.0	2.8	1.0	3.7*	0.8
10	5.4*	0.9	4.5*	1.1	3.3*	1.0	4.2*	0.9

Notes. Impact estimates within 5-km buffers around mines using Callaway & Sant'Anna (2021) estimator for different cluster thresholds. ATT estimates without 95% confidence intervals spanning zero are indicated with *.

Table 1: Cluster threshold variations

D Spillover-robust estimation

A challenge in the empirical strategy is the isolation of one mine's forest loss effect from another proximate mine's spillover effects. Such contamination could bias pre-trend estimates or lead to problems of "double counting", where the forest loss effect of one mine inflates the estimated effect of another. Although important to consider in many settings, DiD designs with spatial spillovers are still poorly understood (Roth et al. 2022). We account for potential spillovers in a ring-estimator framework as outlined in Butts (2023), where control units can be exposed to treatment effects from nearby treated units. The spillover estimation is similar to the two-stage estimator of Gardner (2022), the difference being that the first stage is imputed from observations that are not-yet-treated and, additionally, not-yet-spillover-exposed, i.e. they do not have any active mines located nearby. By including spillover-exposure dummies into the second-stage regression, the treatment effect can be spllit into the spillover deforestation from other mines and the actual deforestation from the mine itself, given that the parallel-trends assumption extends to the spillover area.

Figure 4 shows that the spillover-robust ATT estimates do not differ systematically from the ATTs estimated using the two-stage procedure outlined in Gardner (2022). The spillover effects appear insignificant and relatively constant over time. Therefore, we conclude that spillovers do not impact the validity of our findings.

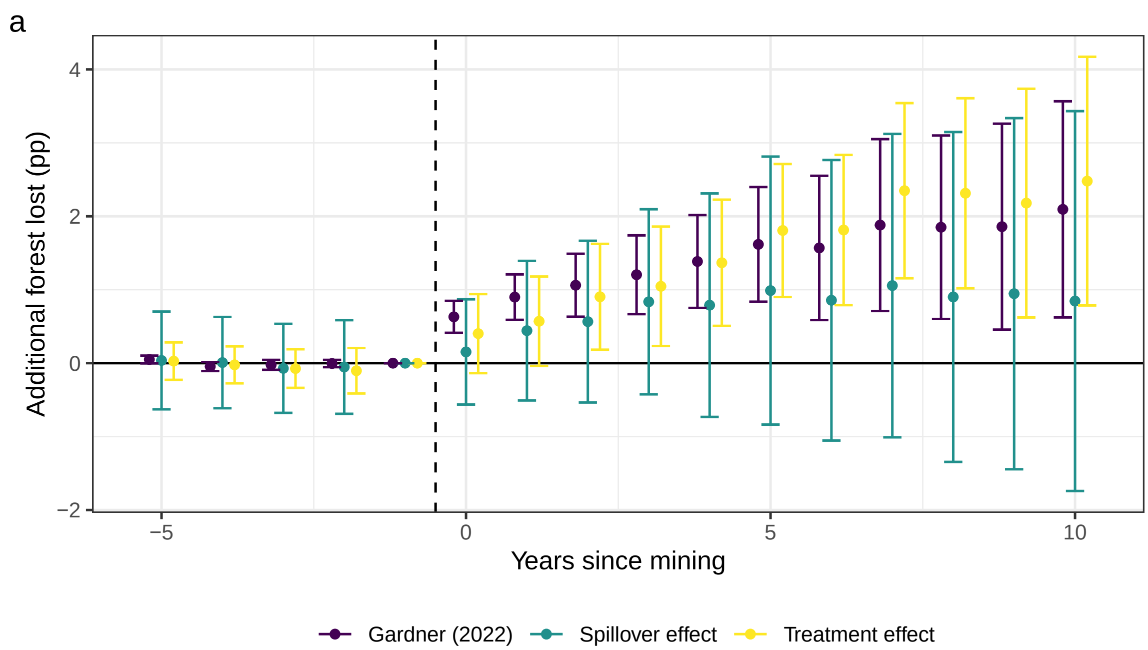

Figure 4: Spillover-robust estimation. "Gardner (2022)" refers to the two-stage estimator proposed in Gardner (2022), while spillover and treatment effect estimates are as suggested in Butts (2023a, 2023b).

E Software

The analysis was conducted in R version 4.3.2. For the data processing, we used the packages *terra* (Hijmans et al. 2023) and *sf* (Pebesma et al. 2023). The statistical analysis relied on the packages *did* (Callaway and Sant'Anna 2022)and *did2s* (Kyle Butts et al. 2023).

References

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2022. "Revisiting Event Study Designs: Robust and Efficient Estimation." http://arxiv.org/abs/2108.12419.

Butts, Kyle. 2023. "Difference-in-Differences Estimation with Spatial Spillovers." http://arxiv.org/abs/2105.03737.

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with multiple time periods." *Journal of Econometrics* 225 (2): 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001.

———. 2022. *Did: Treatment Effects with Multiple Periods and Groups Version 2.1.2 from CRAN*. https://rdrr.io/cran/did/.

Gardner, John. 2022. "Two-stage differences in differences." http://arxiv.org/abs/2207.05943.

Hijmans, Robert J., Roger Bivand, Edzer Pebesma, and Michael D. Sumner. 2023. *Terra Package - RDocumentation*. https://www.rdocumentation.org/packages/terra/versions/1.7-55.

Kyle Butts, John Gardner, Grant McDermott, and Laurent Berge. 2023. *Did2s: Calculate Two-Stage Difference-in-Differences Following... In Did2s: Two-Stage Difference-in-Differences Following Gardner (2021)*. https://rdrr.io/cran/did2s/man/did2s.html.

Masolele, Robert N., Diego Marcos, Veronique De Sy, Itohan-Osa Abu, Jan Verbesselt, Johannes Reiche, and Martin Herold. n.d. "(In Press) Mapping the Diversity of Land Uses Following Deforestation Across Africa." *Scientific Reports*.

Pebesma, Edzer, Roger Bivand, Etienne Racine, Michael Sumner, Ian Cook, Tim Keitt, Robin Lovelace, et al. 2023. *Sf: Simple Features for r*. https://cran.r-project.org/web/packages/sf/index.html.

Pendrill, Florence, Toby A. Gardner, Patrick Meyfroidt, U. Martin Persson, Justin Adams, Tasso Azevedo, Mairon G.Bastos Lima, et al. 2022. "Disentangling the numbers behind agriculture-driven tropical deforestation." *Science* 377 (6611). https://doi.org/10.1126/science.abm9267.

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe. 2022. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." *Journal of Econometrics* 235 (2): 2218–44. https://doi.org/10.1016/j.jeconom.2023.03.008.

Vancutsem, C., F. Achard, J. F. Pekel, G. Vieilledent, S. Carboni, D. Simonetti, J. Gallego, L. E. O. C. Aragão, and R. Nasi. 2021. "Long-term (1990-2019) monitoring of forest cover changes in the humid tropics." *Science Advances* 7 (10): 1–22. https://doi.org/10.1126/sciadv.abe1603.