ChatGPT

Incorporating Land Cover Variables into the Deforestation Panel Model

Relevant Land Cover Variables and Their Roles

In evaluating deforestation reduction in Indonesia (2019–2024), the key land cover variables to consider are:

- **Forest Cover Share** (especially evergreen broadleaf forest): This represents the proportion of land still under forest. It is crucial as it indicates the **remaining forest stock** available to be deforested. The more forest cover an area has, the greater the potential for deforestation – and conversely, once forest is cleared, it cannot be cut again (a non-repeatable outcome) [1]. Forest cover share thus serves as a baseline risk factor for deforestation. Many studies **exclude areas with negligible initial forest** from analysis [2] to avoid trivial zeros in deforestation; this underlines that some baseline forest is needed for deforestation to occur.

- **Cropland (Farmland) Share**: This is the fraction of land already converted to agriculture. It proxies the **pressure and opportunity for agricultural expansion**, a common driver of deforestation. A high cropland share could indicate that an area has historically experienced significant forest-to-farm conversion (leaving less forest remaining), or that it is highly suitable for farming (potentially increasing pressure on any remaining forests). In prior research, expansion of farmland is directly tied to deforestation – for example, **Merkus (2024)** finds that stricter environmental enforcement "reduces conversion rates of forest to farm land," highlighting farmland expansion as a key component of deforestation dynamics [3]. Including cropland share helps capture the **demand for new agricultural land** that might fuel deforestation.

- **Grassland/Open Land Share**: This category often reflects either natural non-forest ecosystems or **previously deforested but not yet cultivated lands** (e.g. scrub, pasture). A sizable grassland share in a district could indicate that forest was cleared in the past and left as degraded land or pasture. Its role is twofold: (1) it may serve as an **alternative land reserve** (i.e. available open land could reduce pressure to clear more forest for agriculture if farmers can use the open lands instead), or (2) it could reflect **low agricultural suitability** (if an area has lots of grassland and little cropland, perhaps the land is not very fertile, which could either discourage further clearing or conversely push expansion into remaining forests if those are seen as more fertile). In practice, grassland share is often complementary to forest and cropland shares (since land cover percentages sum to 100%). Its changes over time might be correlated with deforestation (e.g. if forest is cleared but not immediately farmed, the forest share drops and grassland share rises). This variable can be included to differentiate **types of land conversion** (forest → farms vs. forest → unused open land).

- **Urban/Built-up Share**: This indicates the portion of land that is developed/urban. It tends to be **small and relatively static** in most districts (especially in rural, forested areas) over a short span like 2019–2024. Urban share can proxy for **population density and infrastructure** – areas with higher urban share might have better road access or larger local markets, potentially facilitating deforestation through improved access. On the other hand, a high urban share often coincides with lower forest cover (e.g. in Java or developed areas, forests were cleared long ago). Because urban share changes very little year-to-year, its effect is largely time-invariant (capturing historical development level). This makes it hard to identify in a fixed-effects model (since fixed effects will absorb nearly all of its influence). One might include urban share **as a control for development differences**, but it must be handled carefully due to minimal within-unit variation.

Units and Scaling: It is advisable to use these variables as **shares (percent of land area)** rather than raw area. Using shares normalizes for the fact that administrative units (ADM2 regions) vary in size. For example, 100 km^2 of forest loss in a small district is more significant (a larger share of its land) than 100 km^2 in a very large district. Shares (or fractions) make coefficients easier to interpret in terms of percentage-point changes, and avoid conflating region size with deforestation levels. If area measures are used, one would need to control for total land area to get comparable effects – using shares bypasses this issue.

Integration into the 2SLS Panel Models

Including these land cover variables in the econometric models requires careful consideration of **panel fixed effects** and the **dynamic structure** (lagged dependent variables), especially because some of these land cover shares change slowly over 2019–2024. Below we outline how to incorporate them in (a) a fixed-effects 2SLS model (with potentially a lagged term) and (b) a dynamic panel 2SLS (e.g. Arellano-Bond style) model, and the theoretical reasoning for each approach.

In a Fixed-Effects 2SLS Model (with Lagged Terms)

When using a fixed-effects panel model (e.g. district and year fixed effects) to estimate deforestation drivers, **time-invariant differences** across districts are already controlled. This means the model identifies effects using **within-district changes over time**. For slowly-changing variables like land cover shares, we must ensure we capture their influence on deforestation without causing multicollinearity or absorbing their (small) variation into the fixed effects. Here are recommended approaches:

- **Exclude Near-Static Level Terms:** Avoid including the raw level of a land cover share if it has very little within-unit change. For instance, if evergreen forest cover share only declines from say 60% to 58% over 5 years in a district, a plain fixed-effect estimator may struggle to distinguish this tiny change from noise or common trends. The district fixed effect would capture the fact that district's forest share is ~59%, and the remaining variation (2 percentage points over 5 years) might not be enough for a precise estimate. In such cases, adding the level of forest share as a regressor can lead to imprecise or insignificant coefficients. **Instead, one should incorporate land cover in ways that leverage cross-sectional variation without violating fixed effects.**

- **Use Baseline-Value Interactions (Differential Trends):** A proven strategy (used in prior research) is to interact the initial level of a land cover variable with time indicators. This allows districts with different starting land cover to have different time trends in deforestation. For example, **Burgess et al. (2012)** include "pre-period provincial forest cover interacted with linear trends" in their deforestation regressions [4]. By doing so, they control for the possibility that areas with more forest initially might experience different deforestation trajectories over time than areas with less forest. In our context, we could take each district's **2019 forest cover share** (the baseline) and interact it with either a linear time trend or with year dummies (e.g. baseline_forest_share × 2020, × 2021, …). This introduces a time-varying regressor that represents **the influence of initial forest abundance on deforestation trend**. The coefficient would capture how the passage of time affects deforestation differently in high-forest versus low-forest areas. This approach brings the largely time-invariant forest share into the model **indirectly** (through its interaction with time, which varies each period) [4]. It addresses the non-repeatability issue: regions with more forest to begin with have more to lose and might show higher baseline deforestation rates, but as their forest is depleted, the trend may slow relative to a region that started with modest forest cover.

- **Lagged Land Cover as a Regressor:** Another way to include these variables is by using **lagged values** (which introduces some temporal variation). For example, include the **previous year's forest cover share** (or cropland share) as an explanatory variable for current deforestation. The idea is that deforestation in year t will depend on how much forest was left in year t−1. The more forest remaining last year, the more potential there is for new clearing this year (all else equal) [1]. Conversely, if forest cover was already very low, current deforestation will be naturally constrained. By using the one-period lag, we ensure the regressor is predetermined (forest_{t−1} is not affected by shocks at time t by definition), though it could still be endogenous if, for example, unobserved factors drove both past and current deforestation. If we suspect endogeneity, we can instrument the lagged forest cover (for instance, using forest cover at t−2 or earlier, or other instruments). The **2SLS framework** can accommodate this by treating lagged forest share as an endogenous regressor and using deeper lags as instruments (similar to how one would instrument a lagged dependent variable). This effectively becomes a dynamic model (see next section), but one can implement it in a two-stage least squares setting.

Theoretical note: Including lagged forest cover directly captures the **"risk pool" effect** – as deforestation proceeds, the pool of forest at risk shrinks [1]. This is important for unbiased estimates: if we did not control for remaining forest, a policy that slows deforestation might coincidentally coincide with a period when forest was largely exhausted, confounding interpretation. By controlling for forest_{t−1}, we isolate policy impacts net of the natural slowdown that occurs as forests become scarce.

- **Cropland Share and Other Uses:** Similar logic applies to cropland (farmland) share. We might include **lagged cropland share** to indicate the stage of agricultural expansion. For example, a district with rapidly rising cropland share in t−1 might be experiencing active expansion that could continue to drive deforestation at t. However, note that forest share and cropland share are inversely related – if forest goes down, either cropland or other land categories go up. Including both in the same regression can introduce multicollinearity (they sum to a constant along with other covers). A practical approach is to **include one primary land cover control** and let the fixed effects capture the rest. For instance, we might include forest cover (as the main variable of interest for deforestation) and not include cropland share directly (since if forest % is high, necessarily cropland % is low, given fixed total area). Alternatively, include **cropland share as a control instead of forest share** – this would shift interpretation to how agricultural land proportion affects deforestation. If used, cropland share should also be lagged or interacted with time. For example, one could interact **baseline cropland share** with a time trend, to allow areas that were heavily agricultural at baseline to have a different deforestation trend (possibly lower, since less forest remains, or possibly higher, if agricultural development brings roads that enable new clearing – it's an empirical question). The key is to avoid inserting a mostly time-invariant variable outright; use its lag or an interaction to capture its effect on changes over time.

- **Urban Share and Other Fixed Factors:** Variables like urban share, which barely change, are best treated as **fixed characteristics** absorbed by district fixed effects. If one suspects they influence trends (e.g. more urbanized districts might have had different deforestation trajectories, perhaps deforesting earlier), you could again use an interaction: baseline urban share × time. This would let, say, highly urbanized districts follow a distinct time path in forest loss compared to rural districts. In practice, unless there is a strong hypothesis that urbanization level alters the annual deforestation rate (beyond what forest/farm shares already capture), this may be unnecessary. Most deforestation in the period 2019–2024 is likely in relatively rural locales. Thus, you can reasonably **omit the level of urban share** (letting the fixed effect capture that a district is more or less urbanized), or include a trend interaction if needed for robustness.

- **Avoiding Overfitting with Many Interactions:** If using interactions with year dummies for baseline shares, be mindful of sample size and multicollinearity. With a 2019–2024 panel, one could interact baseline forest share with a single linear trend or with a full set of year indicators (which is more flexible). Burgess et al. (2012) opted for a linear trend interaction for parsimony [4], but they also tried province-specific trends and found results robust. For our purposes, a linear trend × baseline forest share is usually sufficient to capture differential deforestation trajectories between high-forest and low-forest regions. This keeps the model simpler while addressing the core issue.

Summary for FE 2SLS: We propose to include **lagged forest cover share (or baseline-cover-interactions)** as a control in the structural equation. This will account for diminishing forest availability and regional differences in deforestation potential. Cropland share can also be included, ideally through a similar lag or interaction approach, to control for agricultural expansion pressure. By doing so, we base our model on solid theoretical ground: deforestation depends on both **demand-side pressures** (agriculture, development) and **supply-side constraints** (how much forest is left) [1] [3]. Empirically, this mirrors the practices of prior studies: Burgess et al. control for initial forest in trends [4], Merkus focuses on the forest-to-farm conversion mechanism [3], and Assunção et al. (2023) implicitly acknowledge baseline forest by limiting their analysis to areas with sufficient forest cover [2] while using high-frequency alerts to measure enforcement impact. All these suggest that controlling for land cover composition is important for isolating policy effects on deforestation.

In a Dynamic Panel 2SLS (Arellano-Bond style) Model

When we introduce a **dynamic component** – i.e. include a lagged dependent variable (such as last year's deforestation rate) – we typically address it with instruments (since the lagged outcome is endogenous to fixed effects). A common approach is the Arellano-Bond or Arellano-Bover/Blundell-Bond GMM estimators, which use **internal instruments (lags)**. Integrating land cover variables in this context follows slightly different considerations:

- **Differencing and Loss of Level Information:** The first-difference transformation used in dynamic panel estimators will **eliminate time-invariant variables**. If we simply difference the data, a truly time-invariant regressor (like baseline forest cover) disappears (Δbaseline = 0). Even slowly changing variables will lose their level effect: only changes matter after differencing. For example, if forest cover share only changed 2% over the period, the differenced series has values in the ±2% range, and much of the signal (the overall difference between a high-forest and low-forest district) is gone. This means that in a pure difference-GMM, including land cover variables yields little benefit unless those variables exhibit meaningful year-to-year variation. **Grassland and urban shares**, being nearly constant, would effectively drop out. **Forest share** would appear only insofar as it changes due to deforestation (and that change is essentially the negative of the deforestation rate). Including the change in forest cover as a regressor in a deforestation equation would be tautological (deforestation_t is directly related to forest_{t-1} – forest_t). Therefore, we need an alternative way to incorporate land cover in a dynamic panel framework.

- **Use of Lagged Levels in System GMM:** One solution is to use the **System GMM** approach (Blundell-Bond), which combines the differenced equation with a levels equation. In the levels equation, one can include time-invariant or slowly changing covariates under an assumption of exogeneity (or predeterminedness). Practically, we could treat baseline forest cover, or baseline cropland share, as **exogenous variables** that appear in the level equation (with instruments being themselves, assuming they are uncorrelated with the fixed effects after first differencing). This would allow us to estimate coefficients on those baseline shares. In other words, System GMM can recover the influence of persistent covariates by assuming that their correlation with the fixed effect is the same over time and using moment conditions in levels [4]. For example, we might include 2019 forest cover share directly in the model as a regressor (it will be constant for a given district, but in System GMM it can be identified through the level equation). This is a bit advanced, but it's a way to **bring back the level information** that is lost in differencing.

- **Lagged Dependent vs. Lagged Forest Cover:** In a dynamic deforestation model, including the **lagged dependent variable** (e.g. last year's deforestation rate or forest loss) already captures some dynamics. If deforestation is autocorrelated – for instance, an area that had high forest loss last year might continue to have high loss this year due to momentum (new roads, ongoing clearing), or conversely might have lower loss if much of the accessible forest was already cleared – the lagged term will soak up that effect. **However, the lagged deforestation term alone may not fully capture the effect of diminishing forest stock.** It mainly reflects persistence or reversal in the flow of deforestation. To explicitly model the resource constraint, one could include the **lagged level of forest cover** alongside the lagged deforestation. For example, a dynamic specification could be:

$$\mathrm{DeforestedArea}_{it} = \rho\, \mathrm{DeforestedArea}_{i,t-1} + \gamma\, \mathrm{ForestCoverShare}_{i,t-1} + \mathbf{X}_{it}\beta + \text{fixed effects} + \varepsilon_{it},$$

where X_{it} includes other controls (like policy interventions, prices, etc.). Here, γ would capture the effect of having more forest available last period on current deforestation. We expect $ \gamma>0$ (more forest to cut leads to more potential loss), reflecting the **"at-risk forest"** concept [1]. The lagged deforestation term ρ would capture inertia or short-term dynamics (e.g. if $\rho>0$, deforestation tends to continue where it was high; if $\rho<0$, perhaps there's a slowdown after a burst of clearing).

In implementing this, **ForestCoverShare$_{t-1}$ would be treated as an endogenous or predetermined regressor**, instrumented by its earlier lags (similar to the treatment of the lagged dependent variable). Essentially, this is a dynamic panel 2SLS: using instruments like ForestCover${t-2}$ (or further back) to instrument ForestCover$$. This approach is theoretically grounded in the idea that once you control for past deforestation and past forest cover, the remaining error is unexpected shocks, so earlier forest levels (which were set before those shocks) can serve as instruments.

- **Empirical Justification:** Including lagged forest cover in a dynamic model aligns with the intuition from hazard models of deforestation – you can't keep deforesting at high rates once most forest is gone [1]. Empirically, ignoring this can bias results. For example, if a policy started in 2019 and we see deforestation drop by 2024, is it due to the policy or because many districts simply **ran out of forest to cut** by 2024? Controlling for forest stock helps separate these explanations. **Assunção et al. (2023)** effectively addressed this by focusing on areas with intact forest and using high-frequency alerts to measure reductions [5] [2], while **Burgess et al. (2012)** controlled for different baseline forest amounts in trend patterns [4]. Our dynamic specification would explicitly include the remaining forest as a state variable.

- **Cropland in Dynamic Models:** Similar reasoning can apply to cropland share or grassland share, but one must be cautious. If forest cover is in the model, adding cropland share is somewhat redundant (since Cropland${t-1} = 1 - \text{Forest}} - \text{Other{t-1}$ roughly). If there is interest in how agricultural expansion momentum affects deforestation, one could include **lagged cropland share** (instrumented appropriately) instead of lagged forest. This would test if districts that already have a lot of agriculture (or that saw a rise in agriculture last period) have different deforestation rates. For instance, a low Forest$$ (high prior deforestation) often implies high Cropland${t-1}$, which might correlate with less new deforestation simply because little forest remains. Thus, γ on Forest$$ would likely have opposite signs. It may be more straightforward to stick with one of these to avoid multicollinearity in a dynamic context. Given our focus on deforestation, }$ and the corresponding coefficient on Cropland$_{t-1**Forest cover is the more direct choice** to capture the state variable. Cropland share might be left out of the core dynamic specification, or included as a **control for long-run land use** if needed (again, likely as a lagged or baseline interaction term).

- **Time Fixed Effects in Dynamic Model:** Even in the dynamic panel, we would include year fixed effects (common time shocks) to account for any annual fluctuations (e.g. commodity price booms, weather anomalies like El Niño droughts affecting fire incidence, etc.). These year effects will soak up any common trend in deforestation. When we have baseline-share×year interactions in the static model, the analogous effect in a differenced dynamic model is partly absorbed by the combination of year dummies and the lagged variables. If using System GMM, we could still include baseline×year interactions as additional instruments or regressors in the levels equation if we wanted to be very thorough about heterogeneity in trends. However, this can make the GMM instrument count explode and isn't usually necessary unless we detect differing trends that aren't captured by the basic dynamic terms.

Summary for Dynamic 2SLS: We recommend **including the lagged dependent variable (lagged deforestation)** to capture autocorrelation, and **including lagged forest cover share** to capture the resource constraint, both treated as endogenous and instrumented by appropriate lags (classical dynamic panel IV approach). This effectively extends the Arellano-Bond estimator with an additional state variable. The theoretical rationale is to model deforestation as a partially self-regulating process (less forest → less future deforestation) [1], rather than a simple memory-less process. By doing so, the model can distinguish a real policy-induced reduction in deforestation from a mere slow-down because the forest has been largely cleared. Prior empirical work supports this approach: studies often note that after large initial forest losses, deforestation rates drop as the remaining forests are harder to exploit or simply fewer in number [4]. Incorporating lagged forest cover formalizes that insight.

The Proposed Specification and Variables

Based on the above reasoning, our proposal is:

- **Key land cover variable: Forest Cover Share** (evergreen broadleaf forest as % of area). This should enter the model in lagged form (Forest$_{i,t-1}$) to represent remaining forest. It will be included in both the static 2SLS (as an instrumented lag, if we include a lagged dep) and the dynamic panel model (as a predetermined variable). This variable ensures the model accounts for the **diminishing opportunity for deforestation as forests are depleted**.

- **Additional land cover variable: Cropland Share** (% area in cropland). This can be included to control for **agricultural land use pressure**, but carefully. A prudent way is to include it as an initial level interaction with time (in the fixed-effects model) – e.g., include Baseline(2019) cropland share × year trend or year dummies. This would adjust for the possibility that districts with more pre-existing agriculture might have had different deforestation trends (perhaps slower forest loss if much was already cleared, or perhaps faster if agriculture indicates better market access). In the dynamic model, one could include **Cropland$_{t-1}$** alongside Forest$_{t-1}$, but note they are nearly collinear (since one minus the other, aside from grass/urban). An alternative is to include **Cropland expansion** (the change in cropland) as a control: for example, lagged change in cropland area might capture recent agricultural expansion momentum. However, this again is closely tied to recent deforestation (much of cropland change = forest loss if that's the main source of new cropland). So, including cropland is optional and should be done with awareness of multicollinearity. If the focus is on forest conservation, one might present a main specification with just forest cover share, and a robustness check adding cropland share.

- **Grassland/Other Share:** We do not prioritize grassland as a separate regressor in the main model. Its influence is largely the mirror image of cropland in many cases (e.g., if forest goes down and cropland doesn't fully go up, grassland goes up). If there is a hypothesis that type of post-forest land use matters (for example, perhaps deforestation is more likely to continue if previous clearing resulted in productive farms, versus if it created wasteland), one could incorporate **an interaction term** like (forest loss × lagged grassland share) to see if deforestation is more persistent in areas with a lot of idle land. But unless we have specific evidence to test, this may over-complicate the model. Empirically, focusing on forest and cropland covers most of the story – forest cover measures what's left to lose, and cropland share measures what has been gained (the primary alternative use).

- **Urban Share:** We treat this as a control for initial development level, but **not explicitly in the regression** because of its time-invariance. Its impact is accounted for by district fixed effects (each district's unchanging characteristics like population centers, etc., are fixed). If needed, we could add an interaction of **Urban baseline × Year trend** to ensure that highly urban districts aren't driving any spurious trend (e.g. perhaps highly urbanized districts had already slowed deforestation earlier). This is a minor adjustment and likely not pivotal. Given the short panel and the nature of deforestation (concentrated in frontier regions, not in urbanized districts), we can safely omit urban share from the regression without losing explanatory power. The fixed effects and any included covariates (like population or income, if any) would capture similar information.

- **Other Controls:** Although not asked explicitly, it's worth mentioning that we would also include **year fixed effects** and possibly **controls like weather or commodity prices** as used in the literature. For example, **Assunção et al. (2023)** include precipitation, temperature, and agricultural commodity price shocks as controls [6] [7], since these can affect deforestation (e.g., drought can facilitate burning). These controls complement the land cover variables: the land cover variables capture structural differences and constraints, while weather and prices capture annual fluctuations in deforestation incentives. Our land cover variables are stock variables; we'd use them alongside flow variables (like prices or enforcement actions).

By implementing the above, the **2SLS with lag** model will likely take the form of a panel IV regression with district FE, year FE, and an instrumented lag (if using lagged deforestation) or other endogenous variable (e.g. an enforcement measure instrumented by something, as in Assunção's use of cloud cover [5]). The land cover enters as described (baseline interactions or lagged shares). The **dynamic panel 2SLS** (GMM) will use internal instruments for the lagged dependent variable and lagged forest share, ensuring consistency in the presence of fixed effects.

Theoretical and Empirical Rationale

In summary, our approach to include land cover variables is grounded in both theory and prior empirical findings:

- **Theory (Deforestation Dynamics):** Deforestation is a process constrained by natural limits (forest stock) and driven by economic forces (agricultural expansion, infrastructure, etc.). Our model reflects the **constraint** by including forest cover (so that as forest diminishes, the predicted deforestation decreases, ceteris paribus) [1]. It reflects the **drivers** by accounting for cropland (representing agricultural demand for land) [3] and by allowing differences in initial conditions (some regions are more forest-rich, some more agrarian) to influence deforestation paths [4]. This avoids mistaken attribution of effects – for example, without these controls, one might incorrectly conclude a policy was effective simply because deforestation slowed in areas that were running out of forest, or conversely miss a policy effect in areas with abundant forest where deforestation naturally stays high. Including the land cover variables helps isolate policy impacts or other key variables by **conditioning on the appropriate baseline context**.

- **Empirical Precedent:** Our recommendations align with methods in influential studies. **Burgess et al. (2012)**, studying Indonesian deforestation, emphasize controlling for initial forest cover and even exclude observations with essentially no forest [2] to ensure they measure meaningful changes. They introduce interactions of baseline forest with time to capture heterogeneity in trends [4], which we mirror. **Assunção et al. (2023)**, while focusing on Brazil, use an IV strategy (cloud cover as an instrument for enforcement monitoring) and include fixed effects; this implicitly means their variation comes from differences in enforcement exposure **within the same areas over time**, with baseline forest cover differences accounted for by fixed effects and sample restrictions [5] [2]. This strategy was successful in attributing reductions in deforestation to monitoring/enforcement, reinforcing the importance of controlling for confounding factors like forest availability. **Merkus (2024)** demonstrates that when enforcement curtails deforestation, it specifically curtails the conversion of forest into farmland [3] – implying that models of deforestation should track land use transitions. By including cropland share (or observing its change), we ensure our model can capture whether deforestation pressure comes from agricultural conversion (which is often the case in Indonesia, e.g. for oil palm plantations).

- **Small Within-Variation Issue:** We have explicitly addressed the econometric issue that **minimal time variation** in a regressor undermines fixed-effect estimation. Our solution (baseline interactions or lagged terms) ensures that we **leverage cross-sectional variation in a permissible way** under fixed effects. Rather than including a nearly constant variable outright (which would be collinear with the intercept for each district), we use forms that vary over time (like baseline × year) so that estimation is feasible. This approach is commonly recommended in panel data scenarios where key explanatory variables are mostly time-invariant [4]. It lets us retain the valuable information of those variables without falling foul of econometric identification issues.

In conclusion, **we propose to include the evergreen forest cover share (as a proxy for remaining forest) and the cropland share (as a proxy for land-use pressure)** in the deforestation models. Forest cover share should be included in a lagged form or through baseline interactions to account for the resource constraint and to avoid collinearity with fixed effects. Cropland share can be included via baseline interactions or as a lagged variable if appropriate, to control for agricultural expansion influences (with care taken due to its correlation with forest cover). Urban and grassland shares are of lesser importance; their effects are either captured by fixed effects or by the forest/cropland variables (since these three categories partition the land along with minor categories). By structuring the model this way, we ground the analysis in solid theoretical expectations (forest availability and land demand drive deforestation) and draw on the best practices from prior empirical research [4] [3]. This approach will strengthen the credibility of our 2SLS estimates of deforestation reduction effects in Indonesia and ensure we correctly attribute the drivers of any observed changes in forest loss.

Sources:

- Assunção, Juliano, et al. (2023). "DETER-ing Deforestation in the Amazon: Environmental Monitoring and Law Enforcement." AEJ: Applied Economics 15(2): 125–56. (Instrumental variable strategy using satellite alerts; emphasizes effective monitoring curbs deforestation [5] and implicitly focuses on areas with remaining forest [2].)

- Merkus, Erik (2024). "The economic consequences of environmental enforcement: Evidence from an anti-deforestation policy in Brazil." World Development 181: 106646. (Shows that targeted law enforcement **reduces the conversion of forest to farmland**, highlighting the link between deforestation and agricultural land use [3].)

- Burgess, Robin, et al. (2012). "The Political Economy of Deforestation in the Tropics." Quarterly Journal of Economics 127(4): 1707–1754. (Uses satellite-based forest cover data for Indonesia; controls for initial forest cover by interacting it with time trends [4] and excludes areas with minimal baseline forest [2], acknowledging the importance of forest stock in deforestation analysis.)

---

[1] Impact evaluation with nonrepeatable outcomes: The case of forest ...
https://www.sciencedirect.com/science/article/pii/S0095069624000457

[2] [4] OP-QJEC120017 1707..1754
https://economics.mit.edu/sites/default/files/publications/qjefinal.pdf

[3] The economic consequences of environmental enforcement: Evidence from an anti-deforestation policy in Brazil
https://ideas.repec.org/a/eee/wdevel/v181y2024ics0305750x24001165.html

[5] DETER-ing Deforestation in the Amazon: Environmental Monitoring and Law Enforcement - American Economic Association
https://www.aeaweb.org/articles?id=10.1257/app.20200196

[6] [7] aeaweb.org
https://www.aeaweb.org/articles/materials/18285