deforestaion econometrics results outline

Define universal parameters

```
set.seed(930)
b0 = .05
b1 = -0.1
b2 = -0.2
b3 = .16
std_a = .1
std_v = 0.25
years = 3
nobs = 10000
n = 500


cellsize = 15
ppoints = 75
std_p = .2

cpoints = 20
```

1 DGP

The initial data generating process as follows:

We generate the latent variable

$$y^*_{it} = \beta_0 + \beta_1 \mathbb{1}\{D_i = 1\} + \beta_2 \mathbb{1}\{t \geq t_0\} + \beta_3 \mathbb{1}\{D_i = 1\}\mathbb{1}\{t \geq t_0\} + \alpha_i + u_{it},$$

where the mapping from the latent to observed variable y_{it} is

$$y_{it} = \begin{cases} 1 & y^*_{it} > 0 \\ 0 & else \end{cases}$$

The treatment variable is generated according to $D_i \sim bernoulli(.5)$. The period in which the treatment is implemented is denoted t_0. The pixel specific parameter is generated according to $\alpha_i \sim N(0, \sigma_a^2)$ and the error term is generated according to $u_{it} \sim N(0, \sigma_u^2)$\

Note:

$$
\begin{aligned}
ATT =& E[y_{it}(1) - y_{it}(0)|t > t_0, D_i = 1] \\
=& E[y_{it}(1)|t > t_0, D_i = 1] - E[y_{it}(0)|t > t_0, D_i = 1] \\
=& P(y_{it}(1) = 1|t > t_0, D_i = 1) - P(y_{it}(0) = 1|t > t_0, D_i = 1) \\
=& P(y^*_{it}(1) > 0|t > t_0, D_i = 1) - P(y^*_{it}(0) > 0|t > t_0, D_i = 1) \\
=& P(\beta_0 + \beta_1 + \beta_2 + \beta_3 + \alpha_i + u_{it} > 0) - P(\beta_0 + \beta_1 + \beta_2 + \alpha_i + u_{it} > 0) \\
=& P(-\alpha_i - u_{it} < \beta_0 + \beta_1 + \beta_2 + \beta_3) - P(-\alpha_i - u_{it} < \beta_0 + \beta_1 + \beta_2) \\
=& F(\beta_0 + \beta_1 + \beta_2 + \beta_3) - F(\beta_0 + \beta_1 + \beta_2)
\end{aligned}
$$

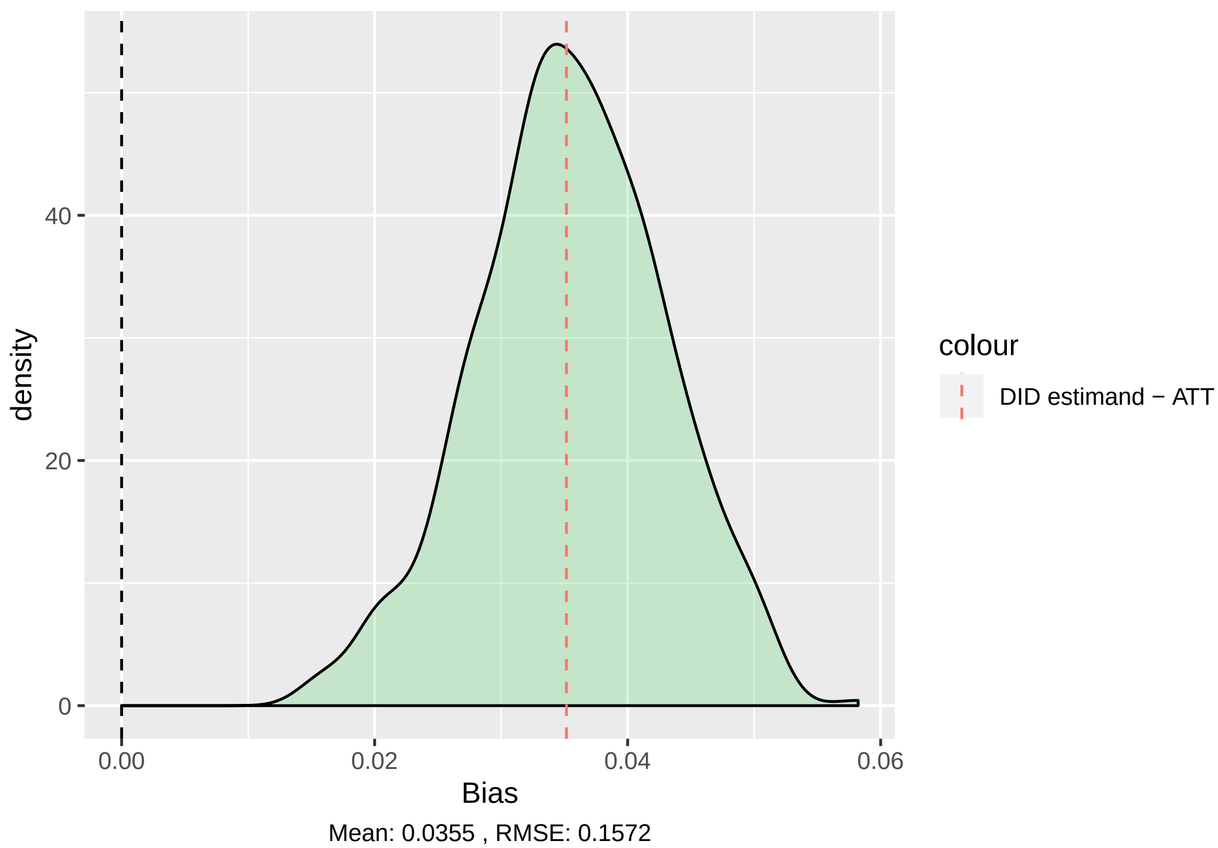
, Where $F()$ is the CDF of a $N(0, \sigma_a^2 + \sigma_u^2)$

2 DID and functional forms

2.1 initial DID estimates

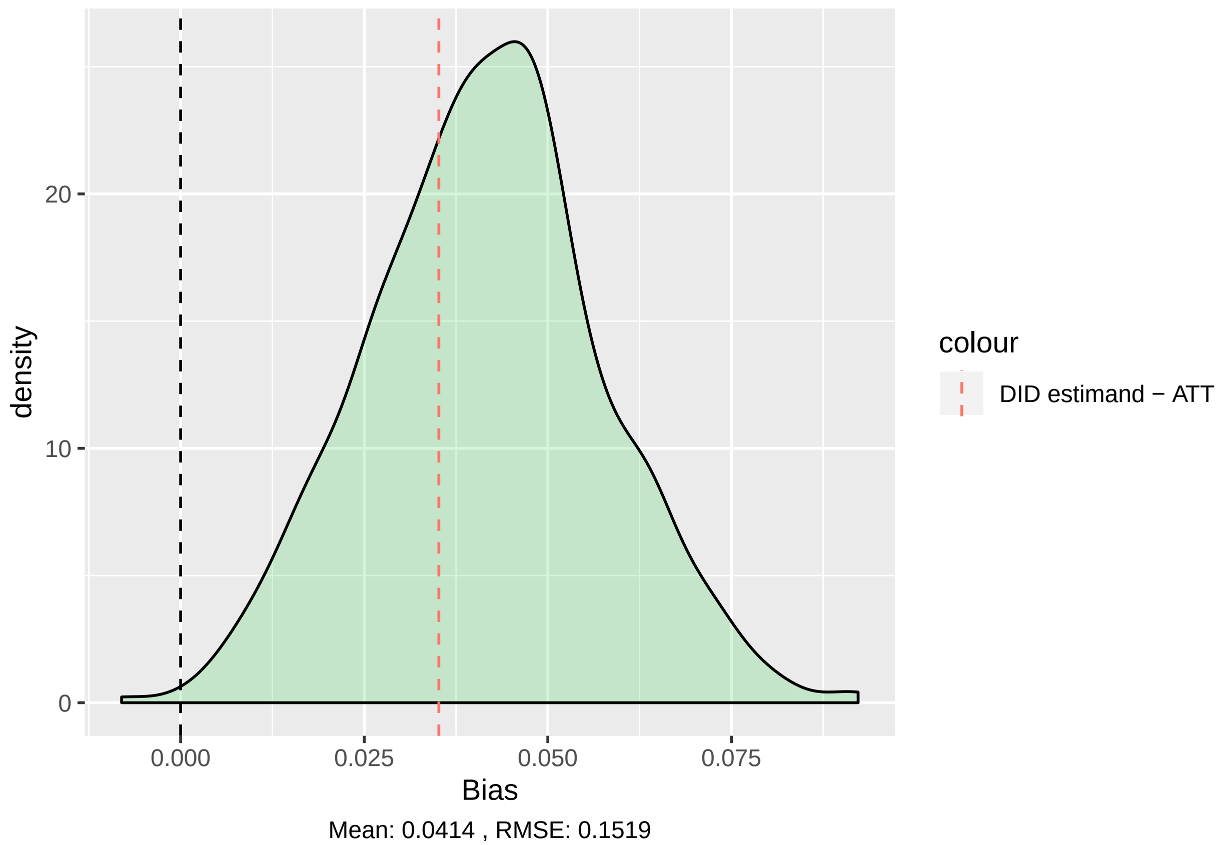
We begin by allowing the outcome to vary between 0 and 1 across time periods. We see a slight bias in the DID estimates of the ATT. This result stems from the fact that the DID estimand does not identify the ATT with the given DGP.

```
DID_y <- quickmontey(n, nobs, years, b0, b1, b2, b3, std_a, std_v, "y")
DID_y$plot
```



We will now drop pixels in the periods after they first become deforested. The DID estimates are still biased here. It appears that the sign of the bias does not change, while the bias worsens slightly.

```
DID_yit <- quickmontey(n, nobs, years, b0, b1, b2, b3, std_a, std_v, "y_it")
DID_yit$plot
```

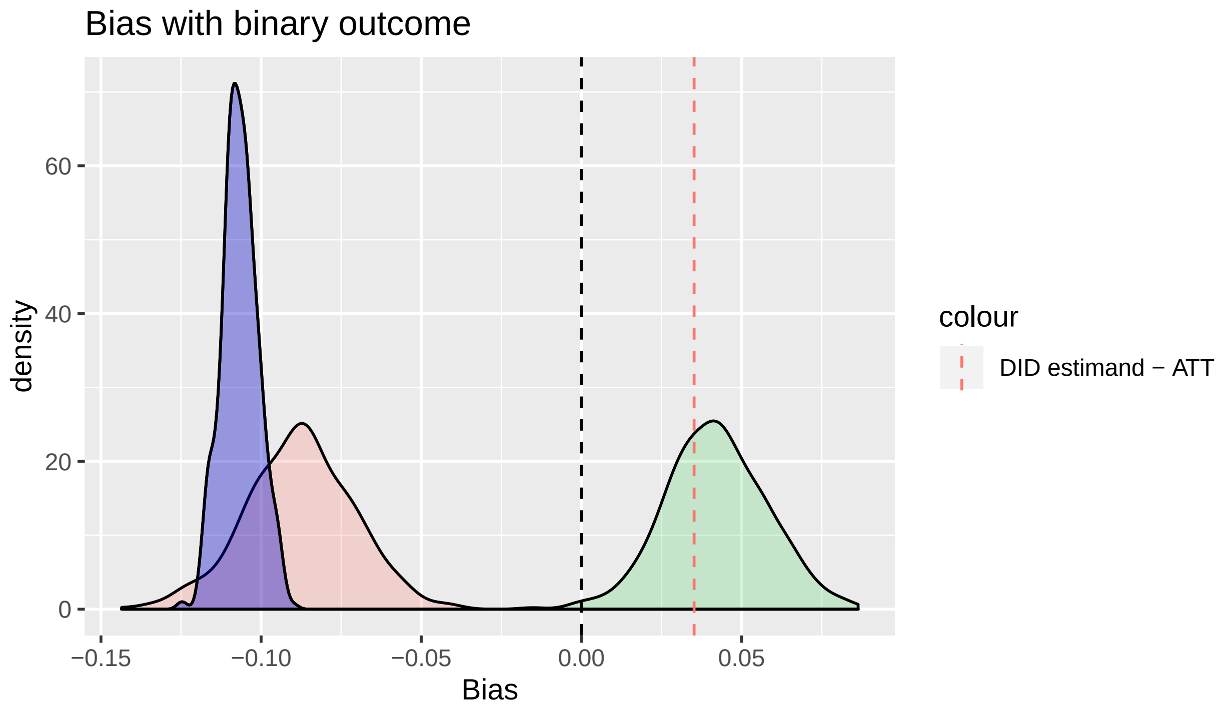


2.2 two-way fixed effects vs. simple DID

We'll now address the use of two-way fixed effects and the dropping of observations in the periods after pixels are first observed as deforested. The magnitude of the bias is smallest using the basic DID. The bias is negative when using the 2way FE dropping deforested obs and when using either the DID or 2way FE keeping obs. Note that the DID and 2way fixed effects estimates are identical when observations are not dropped. As we show in our proof, in the general case, using 2 way fixed effects dropping the observations yields the ATT plus the difference between the treated and untreated groups. The DID on the other hand should identify the ATT.

This may vary in certain cases. If the DID estimand identifies something very far from the ATT, we can imagine that the DID dropping obs may actually generate more bias than the other methods.

```
twowayFE <- binary_coeffdist_fcn(n, nobs, years, b0, b1, b2, b3, std_a, std_v)
twowayFE$plot
```



The mean and RMSE are:
DID dropping obs: 0.042 RMSE: 0.15258
2way FE dropping obs: −0.088 RMSE: 0.14577
DID keeping obs: −0.1066 RMSE: 0.15107
2way FE keeping obs: −0.1066 RMSE: 0.15884

2.3 Two way FE proof

We can show that in the case where the binary outcome is dropped in periods after the outcome is realized as a 1, two-way fixed effects regressions typically do not identify the ATT, but the ATT + the group difference.

pf/

Consider the regression

$$y_{it} = \alpha + \eta d2_t + \tau w_{it} + c_i + u_{it}, \text{ for } t = 1, 2$$

y_{it} is the binary outcome; w_{it} is a dummy equal to 1 when unit i is treated in time t; $d2_t$ is a dummy for the second time period; c_i is an observed effect for unit i\

Jones and Lewis (2015) advise to drop unit i in periods $t+1, ..., T$ when $y_{it} = 1$\

In the two period case, we have

$$y_{i1} = \alpha + \tau w_{i1} + c_i + u_{i1}$$

$$y_{i2} = \begin{cases} \alpha + \eta d2_2 + \tau w_{i2} + c_i + u_{i2} & y_{i1} = 0 \\ NAN & y_{i1} \neq 0 \end{cases}$$

First differencing,

$$y_{i2} - y_{i1} = \begin{cases} \alpha + \eta d2_2 + \tau w_{i2} + c_i + u_{i2} - \alpha - \tau w_{i1} - c_i - u_{i1} & y_{i1} = 0 \\ NAN & y_{i1} \neq 0 \end{cases}$$

Focusing on the first case, where $y_{i1} = 0$

$$\begin{aligned} y_{i2} - y_{i1} &= \alpha + \eta d2_2 + \tau w_{i2} + c_i + u_{i2} - \alpha - \tau w_{i1} - c_i - u_{i1} \\ &= \alpha + \eta d2_2 + \tau w_{i2} + c_i + u_{i2} - \alpha - c_i - u_{i1} \\ &= \eta d2_2 + \tau w_{i2} + u_{i2} - \tau w_{i1} - u_{i1} \\ &= \eta + \tau w_{i2} + \Delta u_i \end{aligned}$$

Coming back to the general expression,

$$y_{i2} - y_{i1} = \begin{cases} \eta + \tau w_{i2} + \Delta u_i & y_{i1} = 0 \\ NAN & y_{i1} \neq 0 \end{cases}$$

(NAN - $(\alpha + \tau w_{i1} + c_i + u_{i1})$ is NAN)\

We have that with binary w_{i2}

$$\hat{\tau} = \frac{1}{n_{ST}} \sum_{i:w_i=1} y_{i2} - \frac{1}{n_{ST}} \sum_{i:w_i=1} y_{i1} - (\frac{1}{n_{SU}} \sum_{i:w_i=0} y_{i2} - \frac{1}{n_{SU}} \sum_{i:w_i=0} y_{i1})$$

, where n_{ST} and n_{SU} are the number of surviving treated and untreated units such that $y_{i1} = 0$, respectively.

Note: Since this captures where $y_{i1} = 0$, we have $\frac{1}{n_{ST}} \sum_{i:w_i=1} y_{i1} = 0$ and $\frac{1}{n_{SU}} \sum_{i:w_i=0} y_{i1} = 0$ \

Then,

$$\begin{aligned} \hat{\tau} &= \frac{1}{n_{ST}} \sum_{i:w_i=1} y_{i2} - 0 - (\frac{1}{n_{SU}} \sum_{i:w_i=0} y_{i2} - 0) \\ &= \frac{1}{n_{ST}} \sum_{i:w_i=1} y_{i2}(1) - \frac{1}{n_{SU}} \sum_{i:w_i=0} y_{i2}(0) \\ &\text{(whether we see treated or untreated outcome)} \\ &= \frac{1}{n_{ST}} \sum_{i:w_i=1} y_{i2}(1) - y_{i2}(0) \\ +&\frac{1}{n_{ST}} \sum_{i:w_i=1} y_{i2}(0) - \frac{1}{n_{SU}} \sum_{i:w_i=0} y_{i2}(0) \\ &\text{(adding and subtracting } \frac{1}{n_{ST}} \sum_{i:w_i=1} y_{i2}(0)) \\ &= ATT + Diff \end{aligned}$$

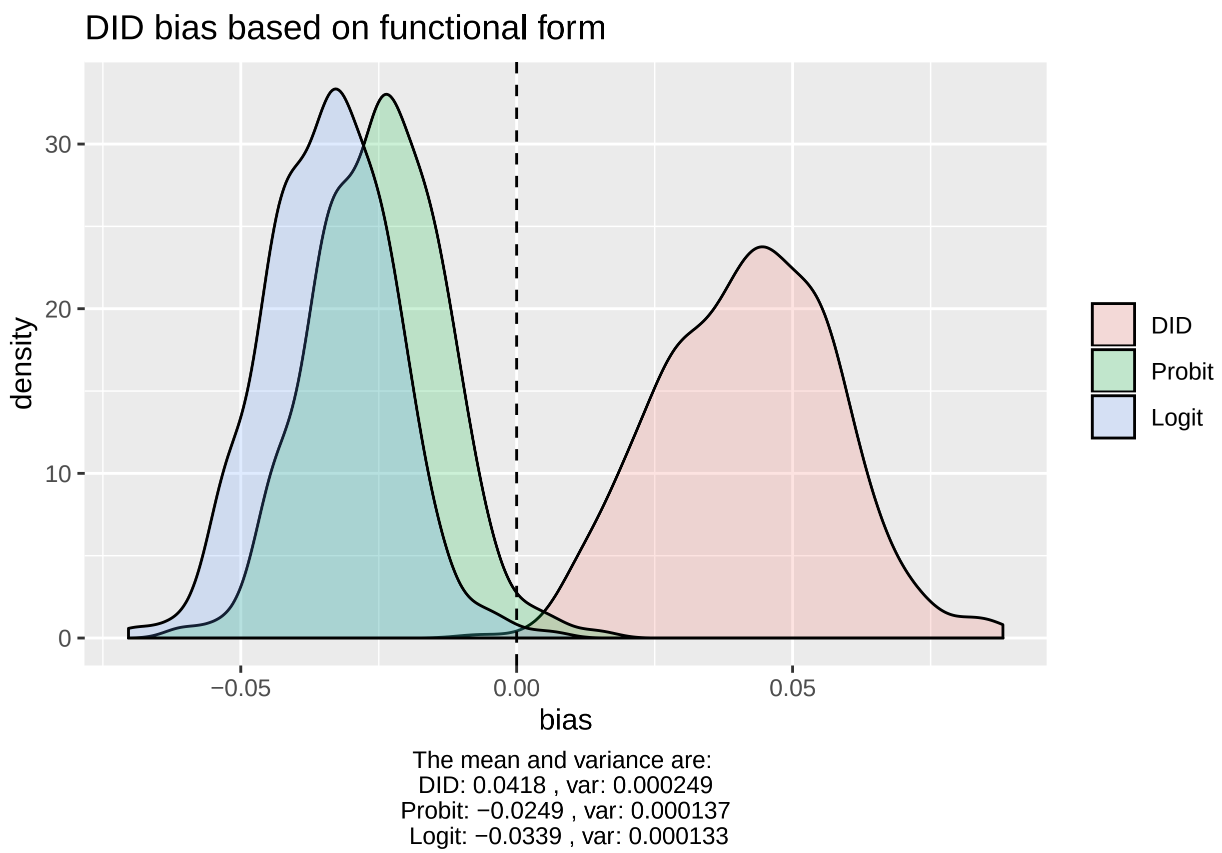
■

2.4 functional form test (logit, probit, vs. DID)

```
functionalform <- funcform(n, nobs, years, b0, b1, b2, b3, std_a, std_v)
```

```
## No id variables; using all as measure variables
```

```
functionalform$plot
```



The above plot shows the DID bias depending on functional form decisions. Interestingly, in our case the magnitude of the bias is fairly similar using each of the three methods. The probit seems to outperform the logit slightly. The sign does differs between the probit/logit and DID estimates however.

3 Aggregating pixels

3.1 various outcomes wheen aggregating

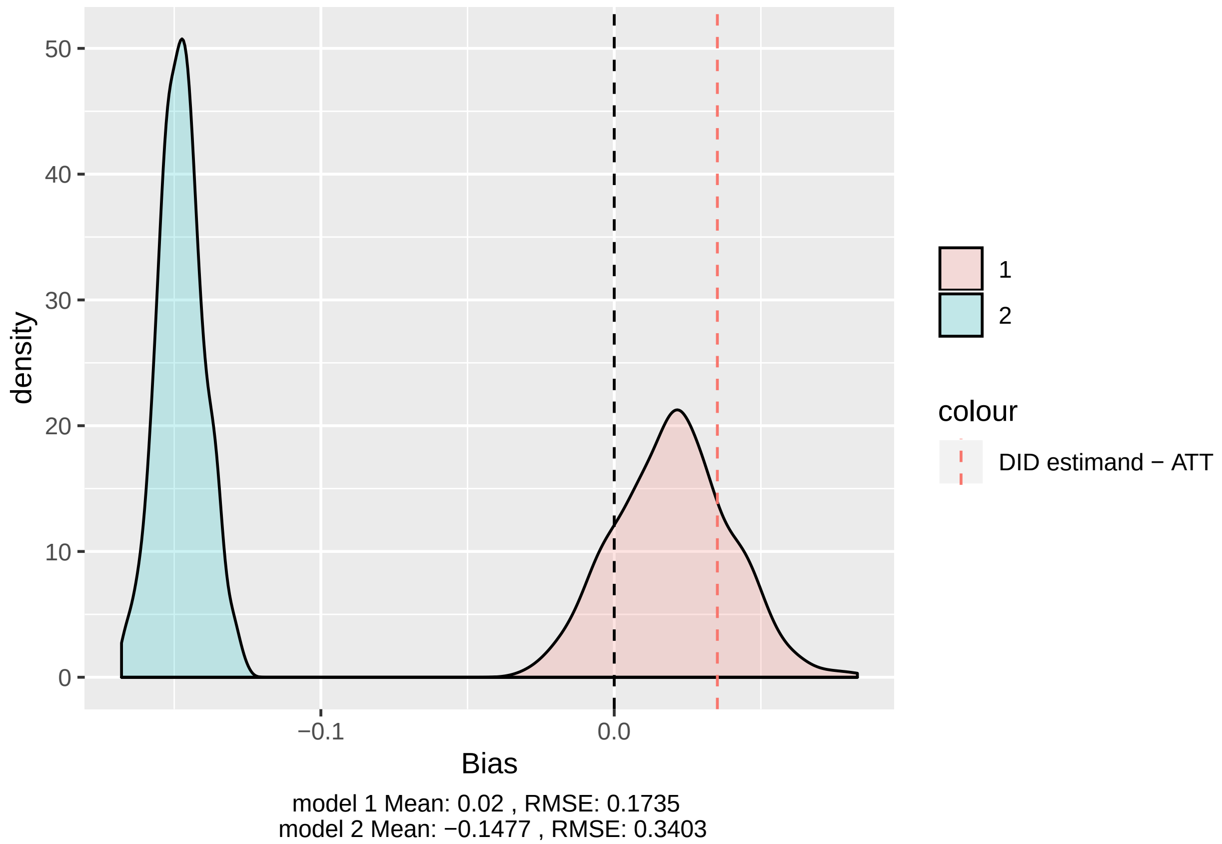
We now aggregate to the grid level. Below, we see the bias introduced by using different outcome variables and specifications commonly used in the literature. (Need a table here to show the different outcomes and will want to cite different papers that use these outcomes. Currently only two outcomes are being used in this function)

number	outcome	additional covariates	model type	paper
1	$\frac{F_{it-1}-F_{it}}{F_{it-1}}$		Two-way FE	
2	$\frac{F_{i0}-F_{it}}{F_{i0}}$		Two-way FE	

```r
outcomes <- outcome_comparison(n, nobs, years, b0, b1, b2, b3, std_a, std_v, cellsize)
```

```
## No id variables; using all as measure variables
```

```r
outcomes$plot
```



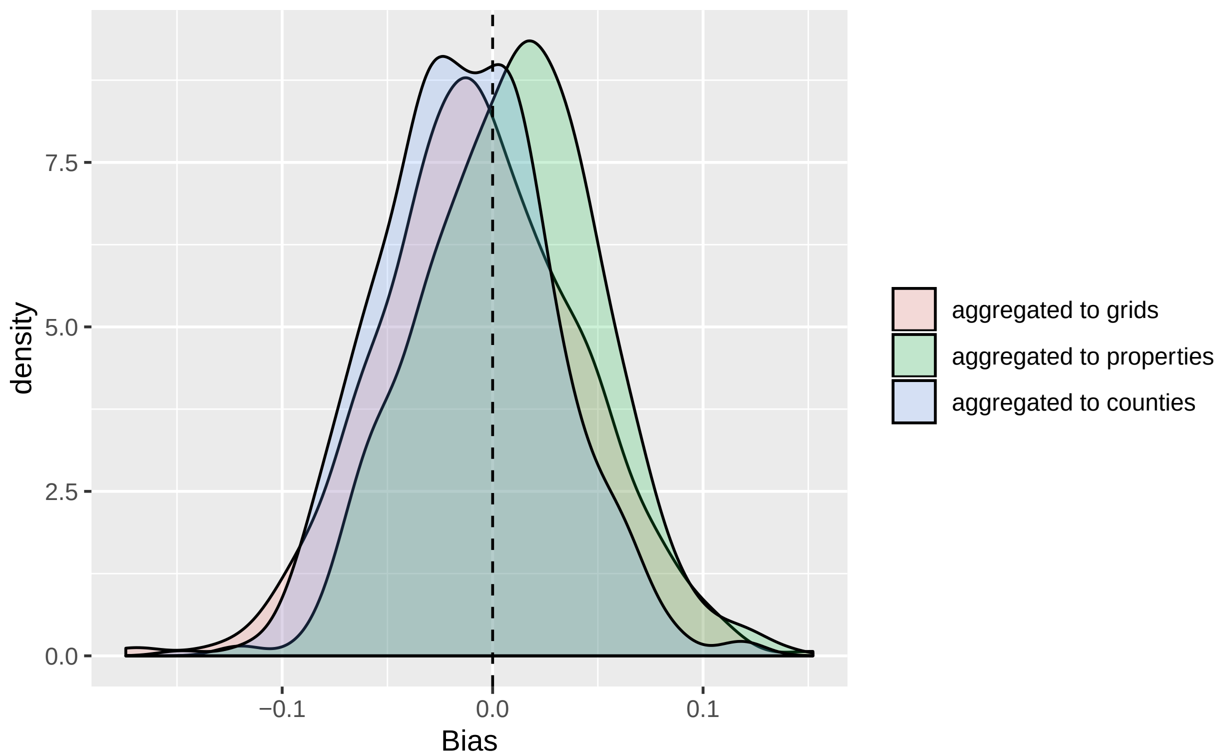
3.2 aggregating to county, property, and grid levels

Moving forward, we use the first outcome from above, as it generated the least bias. We consider three possible aggregation methods: county, property, and grid level in order to look at the distribution of the estimates. At this point we also introduce property level perturbations into the DGP. The property level aggregation leads to the least bias in this case, however the distributions appear to be relatively similar for all three levels of aggregation.

```r
countycoverage <- clustercover(n, nobs, years, b0, b1, b2, b3, std_a, std_v, std_p, cellsize, p
```

```
## No id variables; using all as measure variables
```

```r
countycoverage$plot
```



Mean grid: −0.0062 , RMSE: 0.1788
Mean property: 0.0095 , RMSE: 0.1621
Mean county: −0.0126 , RMSE: 0.1833

3.3 Coverage based on aggregation and standard error

We consider three possible aggregation methods: county, property, and grid level. We then compute the coverage probability of the ATT with a 95% CI. We first cluster standard errors at the group level and then, simply use White standard errors for comparison.

Property level aggregation also leads to roughly the expected coverage, while the county and grid level aggregation results in under coverage. It is unclear how much of the issue is related to the bias rather than the standard errors themselves. The coverage seems to be the same whether we cluster the standard errors at the group level or simply use White standard errors.

level of aggregation	std. errors	coverage probability
grid	clustered at grid	0.914
property	clustered at property	0.958
county	clustered at county	0.902
grid	white	0.914
property	white	0.958
county	white	0.902

3.4 Weighting the regression

Interestingly, we note that upon weighting the regression aggregated to the property level, the distribution is significantly wider. This also appears to happen with the regression aggregated to the county level, but not nearly to the same extent. Weighting the regression aggregated to the grid level has almost no impact. This makes intuitive sense, since the grids are all of equal size.

Weighting the regression results in less bias at the property level, but the difference is negligible at the county and grid level. We see a larger impact of weighting the regressions in the property and county cases, because there is more area heterogeneity across units.

```
weighting <- weightingarea(n, nobs, years, b0, b1, b2, b3, std_a, std_v, std_p, cellsize, ppoi
```

```
## No id variables; using all as measure variables
```

```
weighting$plot
```