HW3
Labor Economics

Chia-wei, Chen*

October 16, 2022

1 Normalization of the Selection Equation

1.1 Distribution of $F_U(U)$

We consider the CDF of the CDF, that is,

> How much of the U's will the CDF of U be less than y?

Apparently this is exactly the definition of a CDF, hence this number is trivially y, and therefore we know

$$\Pr(F_U(U) \leq y) = y$$

, which is a uniform distribution between $[0, 1]$

1.2 $\mathscr{P}(z) = F_U(v(z))$ [1]

The original definition of selection is

$$D = \mathbb{1}(U \leq v(Z)) = \mathbb{1}(\tilde{U} \leq F_U(v(Z)))$$

where $\tilde{U} \equiv F_U(U)$. This holds because CDF is a increasing function.

Therefore inside the propensity score we substitute $D = 1$ with the condition

$$F_U(U) \leq F_U(v(Z))$$

*R10323045

[1] I denote the propensity score as $\mathscr{P}$.

Hence

$$\mathscr{P}(z) = \Pr(D = 1 \mid Z = z) = \Pr(F_U(U) \leq F_U(v(Z)) \mid Z = z) = \Pr(F_U(U) \leq F_U(v(z)))$$

Because $F_U(U)$ is uniformly distributed, we conclude that

$$\mathscr{P}(z) = F_U(v(z))$$

1.3 Show that $D = \mathbb{1}\{\tilde{U} \leq \mathscr{P}(Z)\}$

A direct substitution should do the work.

$$D = \mathbb{1}(U \leq v(Z)) = \mathbb{1}(\tilde{U} \leq F_U(v(Z))) = \mathbb{1}(\tilde{U} \leq p(Z))$$

2 Derivation of the Weights for LATE

$$\text{LATE}_{z'}^{z} = \frac{E[Y \mid Z = z] - E\left[Y \mid z = z'\right]}{E[D \mid Z = z] - E\left[D \mid z = z'\right]} = \int_0^1 \text{MTE}(u) \times \frac{1\left\{u \in \left[p\left(z'\right), p(z)\right]\right\}}{p(z) - p\left(z'\right)} du$$

2.1

Proof. First, note that $\mathscr{P}(z) = \Pr(D = 1 \mid Z = z)$, which is the proportion of observations that gets the treatment given the state of instrument z, which will be $\mathbb{E}(D \mid Z = z)$.

Accordingly, $\mathbb{E}(D \mid Z = z) - \mathbb{E}(D \mid Z = z') = \mathscr{P}(z) - \mathscr{P}(z')$, thus corresponds to the denominator of the weight. $\square$

2.2

Show that $\mathbb{E}[Y \mid Z = z] = \mathbb{E}[Y_1 \mid U \leq \mathscr{P}(Z)]\mathscr{P}(z) + \mathbb{E}[Y_0 \mid U > \mathscr{P}(Z)](1 - \mathscr{P}(z))$

Proof. We start by expression y in the potential outcome framework

$$\begin{aligned}
\mathbb{E}[Y \mid Z = z] &= \mathbb{E}[Y_1 D + Y_0(1 - D) \mid Z = z] \\
&= \mathbb{E}[Y_1 D \mid Z = z] + \mathbb{E}[Y_0(1 - D) \mid Z = z]
\end{aligned}$$

The probability of $D = 1$ given z is the propensity score $\mathscr{P}(z) = \Pr(D = 1 \mid Z = z)$, and $D = 1$ implies $\tilde{U} \leq \mathscr{P}(Z)$. Together we get

$$= \mathbb{E}[Y_1 \mid \tilde{U} \leq \mathscr{P}(z)]\mathscr{P}(z) + \mathbb{E}[Y_0 \mid \tilde{U} > \mathscr{P}(z)](1 - \mathscr{P}(z)) \tag{1}$$

Where $1-\mathscr{P}(z)$ is the probability of not choosing $D = 1$, which is the probability of choosing $D = 0$ □

2.3

Show that

$$\mathbb{E}[Y \mid Z = z] = \int_0^{\mathscr{P}(z)} \mathbb{E}[Y_1 \mid \tilde{U} = u]du + \int_{\mathscr{P}(z)}^1 \mathbb{E}[Y_0 \mid \tilde{U} = u]du$$

Proof. Let us start with $\mathbb{E}[Y_1 \mid \tilde{U} \leq \mathscr{P}(z)]$ in Eq. (1). The condition is a range, hence we expand it into

$$\mathbb{E}[Y_1 \mid \tilde{U} \leq \mathscr{P}(z)] = \int_0^1 \frac{\mathbb{E}[Y_1 \mid \tilde{U} = u]}{\Pr(\tilde{U} \leq \mathscr{P}(z))} \mathbb{1}\{u \leq \mathscr{P}(z)\} dF_{\tilde{U}}(u)$$

Since $\tilde{U} \sim \text{Unif}[0, 1]$, $\Pr(\tilde{U} \leq \mathscr{P}(z)) = \mathscr{P}(z)$, and $dF_{\tilde{U}}(u) = 1du$. The above term simplifies to

$$\frac{1}{\mathscr{P}(z)} \int_0^{\mathscr{P}(z)} \mathbb{E}[Y_1 \mid \tilde{U} = u]du$$

Similarly,

$$\mathbb{E}[Y_0 \mid \tilde{U} > \mathscr{P}(z)] = \frac{1}{1 - \mathscr{P}(z)} \int_{\mathscr{P}(z)}^1 \mathbb{E}[Y_0 \mid \tilde{U} = u]du$$

Substitute into Eq. (1), we get

$$\mathbb{E}[Y \mid Z = z] = \int_0^{\mathscr{P}(z)} \mathbb{E}[Y_1 \mid \tilde{U} = u]du + \int_{\mathscr{P}(z)}^1 \mathbb{E}[Y_0 \mid \tilde{U} = u]du$$

□

2.4

Proof.

$$\int_0^{\mathscr{P}(z)} \mathbb{E}[Y_1 \mid \tilde{U} = u]du + \int_{\mathscr{P}(z)}^1 \mathbb{E}[Y_0 \mid \tilde{U} = u]du - \int_0^{\mathscr{P}(z')} \mathbb{E}[Y_1 \mid \tilde{U} = u]du - \int_{\mathscr{P}(z')}^1 \mathbb{E}[Y_0 \mid \tilde{U} = u]du$$
$$= \int_{p(z')}^{p(z)} \mathbb{E}[Y_1 \mid \tilde{U} = u]du - \int_{p(z')}^{p(z)} \mathbb{E}[Y_0 \mid \tilde{U} = u]du$$
$$= \int_{p(z')}^{p(z)} \mathbb{E}[Y_1 - Y_0 \mid \tilde{U} = u]du$$
$$= \int_{p(z')}^{p(z)} \mathrm{MTE}(u)du$$

□

3 Policy Relevance Treatment Effect

3.1 ATE

The ATE measures

$$\mathbb{E}[Y_1 - Y_0]$$

It is the expected difference of future average earning on whether one attends college or not.

3.2 ATT

The ATT measures

$$\mathbb{E}[Y_1 - Y_0 \mid D = 1]$$

It is the expected difference of future average earning for people that attended college. This is a what if question: *What will be a bachelors' future earning if he doesn't go to college?*

3.3 PRTE

Although the treatment effect mentioned above tells us some aspects of potential outcome, it is not really useful for policy makers. For example, given a ATE, a consultant tell the policy makers *This is the average different of earning, but I can't tell you whether changing the tuition makes any different.* Similarly, the ATT tell the policy maker *I don't know whether this policy make people tend to attend college more or not, but if they already attend college, this will be how much they earn.*

The policy relevant treatment effect, on the other hand, considers together the difference in the outcome we are interested in (Y) as well as the change in the decision making process according to a policy driven channel ($Z*$ changing D).

With PRTE, we can estimate the effect of conducting a policy.

3.4 Relationship with LATE

The local average treatment effect is defined as

$$\text{LATE} = \mathbb{E}[Y_1 - Y - 0 \mid D_z = z, D_{z*} = z*]$$

and its estimation

$$\widehat{LATE} = \frac{\mathbb{E}[Y \mid Z = z*] - \mathbb{E}[Y \mid Z = z]}{\mathbb{E}[D \mid Z = z*] - \mathbb{E}[D \mid Z = z]}$$

Notice that the only different with PRTE is the condition. Requiring $\mathbb{E}(Y*) = \mathbb{E}[Y \mid Z = z*]$ and vice versa is the key for the two to match. Intuitively, the PRTE will be equivalent to LATE if the samples that we consider are the ones that given a change in policy, it will definitely react to it.

4 Arellano-Bond

$$Y_{it} = \rho Y_{it-1} + \delta_i + \epsilon_{it}$$

with $Cov(\epsilon_{it}, Y_{is}) = 0 \quad \forall s \leq t - 1$

4.1 Show the inconsistency of FE estimator

We first demean the entire model

$$Y_{it} - \bar{Y}_i = \rho(Y_{it-1} - \bar{Y}_{i,-1}) + (\epsilon_{it} - \bar{\epsilon}_i)$$

This is an OLS with no constant term. Ideally we can extract ρ by regressing $Y_{it} - \bar{Y}_i$ on $(Y_{it-1} - \bar{Y}_{i,-1})$, but there exist correlation between the error term and the regressor, therefore it is doomed to be inconsistent (if T is not big enough).

Proof.

$$
\begin{aligned}
&Cov((Y_{it-1} - \bar{Y}_{i,-1}), (\epsilon_{it} - \bar{\epsilon_i})) \\
&= Cov(Y_{it-1}, \epsilon_{it}) - Cov(Y_{it-1}, \bar{\epsilon_i}) - Cov(\bar{Y}_{i,-1}, \epsilon_{it}) + Cov(\bar{Y}_{i,-1}, \bar{\epsilon_i})
\end{aligned} \tag{2}
$$

Note that

$$
\begin{aligned}
Cov(Y_{it-1}, \bar{\epsilon_i}) &= Cov(Y_{it-1}, \frac{1}{T}\sum_0^T \epsilon_{it}) \\
&= \frac{1}{T} Cov(Y_{it-1}, \epsilon_{it-1}) \neq 0
\end{aligned}
$$

The last equation comes from the assumption that $Cov(\epsilon_{it}, Y_{is}) = 0 \quad \forall s \leq t-1$
Other terms in Eq. (2) are zero according to the assumption as well.

Therefore we prove that as long as T is finite, the fixed effect estimation for ρ is inconsistent.

□

.

4.2 Taking first difference

$$
\begin{array}{rcl}
Y_{it} &=& \rho Y_{it-1} + \delta_i + \epsilon_{it} \\
-)Y_{it-1} &=& \rho Y_{it-2} + \delta_i + \epsilon_{it-1} \\
\hline
Y_{it} - Y_{it-1} &=& \rho(Y_{it-1} - Y_{it-2}) + (\epsilon_{it} - \epsilon_{it-1})
\end{array}
$$

4.3 OLS on first difference

The correlation between the error term and the regressor still remains.

Proof.

$$
\begin{aligned}
&Cov((Y_{it-1} - Y_{it-2}), (\epsilon_{it} - \epsilon_{it-1})) = ... \\
&= Cov(Y_{it-1}, \epsilon_{it-1}) \neq 0
\end{aligned}
$$

□

We can, however, choose an instrumental variable such that it correlates with y_{it-1} but not y_{it}. The second lag term will Y_{it-2} be a good choice.

5 TWFE is Biased

5.1 Plot the concept

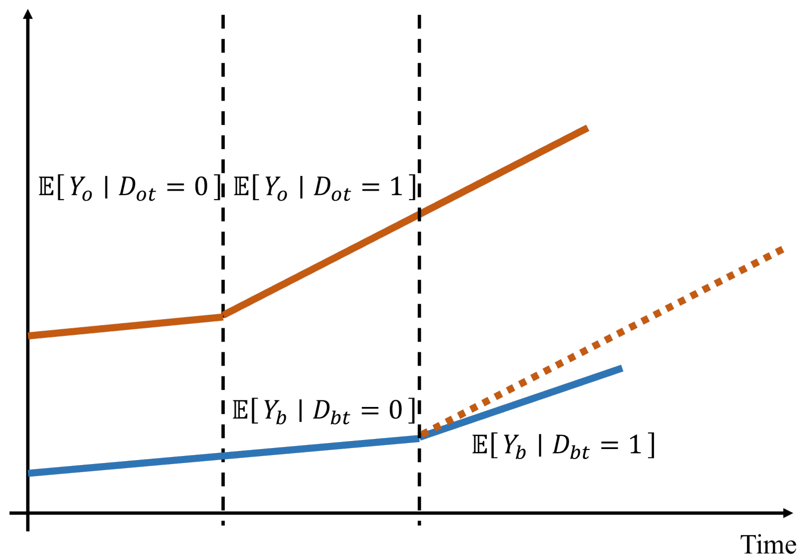

Figure 1: A failure of simply using the two way fixed effect on staggered treatment effect

See figure 1. The purpose of DiD is to extract the trend of the treated in a way that the counterfactual outcome can be obtained. The TWFE, however, takes the treated (the orange line after T_1) as the controlled group, causing the DiD on the blue cohort to actually look to have a downward effect.

5.2 Simulation

Referring to an excellent article written by Callaway and Sant'Anna (2022), I replicated a simplified version of the concept [2].

The data generating process is as follow

$$Y_{i,t,g} = (2020 - g) + \alpha_i + \alpha_{t,g} + \tau_{i,t} + \epsilon_{i,t}$$

where i is the unit, t is the year, and g is the group. Each unit belongs to a state. There are 40 states in this data generating process, each randomly assigned

[2] https://cran.r-project.org/web/packages/did/vignettes/TWFE.html

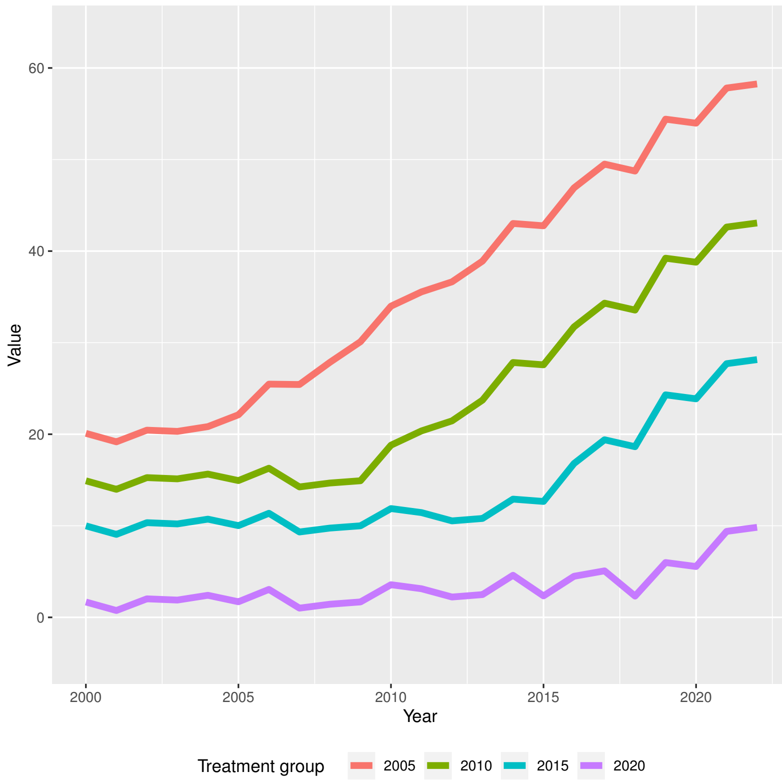

Figure 2: Data for Monte-Carlo simulation

to one of the treated-year in the group $\{2005, 2010, 2015, 2020\}$. Therefore each unit corresponds to one of the group, and hence is treated in one of the year above. The fixed effects are defined as follow

$$\begin{aligned}
\alpha_i &\sim \mathcal{N}(0.2state, 1)\\
\alpha_t &\sim 0.1(t-g) + \epsilon_t^{FE}\\
\epsilon_t^{FE} &\sim \mathcal{N}(0,1)\\
\tau_{i,t} &= \mu(t-g+1)\mathbb{1}\{t \geq g\}\\
\epsilon_{i,t} &\sim \mathcal{N}(0, \frac{1}{4})
\end{aligned}$$

The time series data is shown in figure 2. Here the treatment effect μ is set to be 2.

Note that α_t is a cohort specific parallel time-trend, the expected value of time fixed effect for each group will then be set to 0 when they are treated.

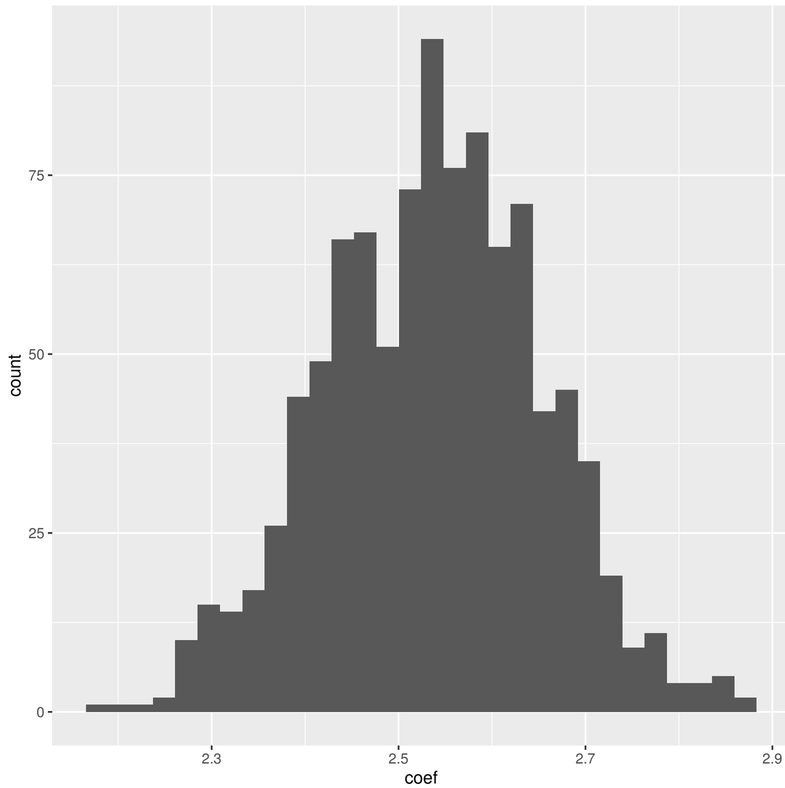

Figure 3: Monte Carlo simulation using the twoway fixed effect estimation with $N = 1000$

We then estimate the following TWFE

$$Y_{i,t} = \alpha_i + \alpha_t + D_{it}\beta_{TWFE} + \epsilon_{it}$$

The result is shown in figure 3

We test the hypothesis

$$H_0 : \hat{\mu} = 2$$

using a two tail t-test. The t-value is 147.84, and the confidence interval is $[2.529415, 2.543658]$. We reject the null hypothesis and conclude that the TWFE estimator causes a bias on this scenario.

5.3 The `did` Library

Using the `did` package contributed by Callaway and Sant'Anna (2021), we get the following result:

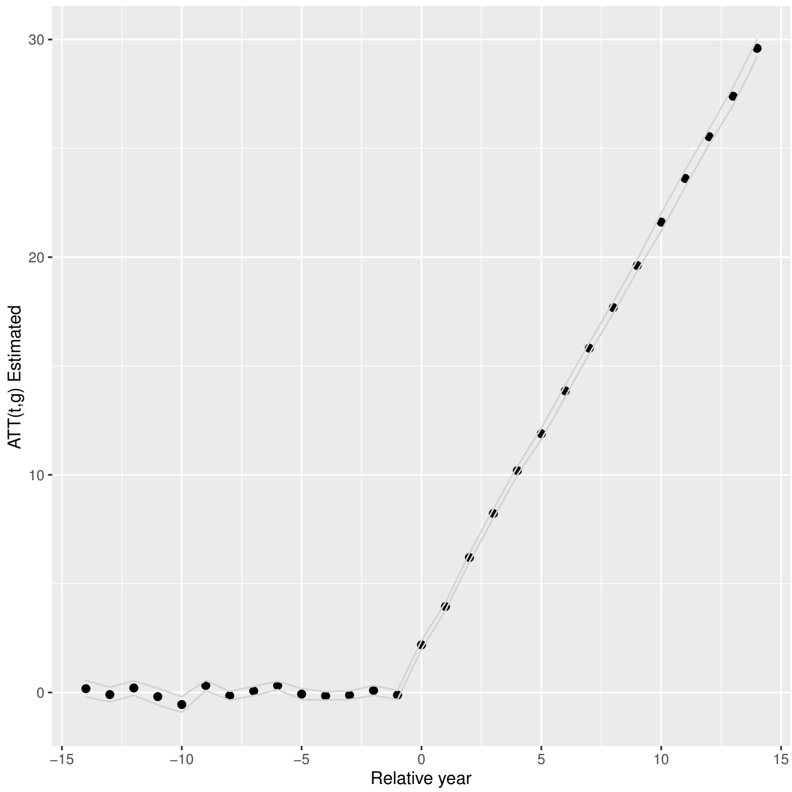

Figure 4: $ATT(g,t)$ estimated by the `did` package. The ribbon represents $\pm 1 s.e.$

The slope is roughly 2, which matches the setting of our data generating process.

5.4 Code

```r
library(tidyverse)
library(plm)
library(did)
mu = 2 # the real treatment effect

nobs = 100
year_range =  2000:2022
nstates = 40
years_treat = c(2005, 2010, 2015, 2020)


############################
# Data generating process
#############################

DGP = function(){
  alpha_i = tibble(
    unit = 1:nobs,
    state = sample(1:nstates, nobs, replace = T),
    unit_fe = rnorm(nobs, state/5, 1)
  )

  alpha_t = tibble(
    year = year_range,
    year_fe = rnorm(length(year), 0, 1)
  )

  # out of 40 states, there are some that belongs to some treatment group
  state_treatments = tibble(
    state = sample(1:nstates, nstates, replace = F),  # shuffle
    year_of_treatment = rep(years_treat, nstates/length(years_treat))
  )


  ## combine all together.
  ## first create a big table of n_obs * n_years rows
  expand.grid(unit = 1:nobs, year = year_range) %>%
    left_join(., alpha_i, by="unit") %>%
    left_join(., alpha_t, by="year") %>%
    left_join(., state_treatments, by="state") %>%
    mutate(
      treat = ifelse(year >= year_of_treatment, 1,0),
```

```
      tau = ifelse(treat==1, mu, 0)*(year - year_of_treatment +1),
      error = rnorm(n(),0,1),
      year_fe = year_fe + 0.1 * (year - year_of_treatment),
      y_value = (2022 - year_of_treatment) + unit_fe + year_fe + tau
    + error
    )
}


##############################
# Visualizing the DGE
##############################

df = DGP()
plot1 <- df %>%
  ggplot(aes(x = year, y = y_value, group = unit)) +
  geom_line(alpha = 1/8, color = "grey") +
  geom_line(data = df %>%
              group_by(year_of_treatment, year) %>%
              summarize(y_value = mean(y_value)),
            aes(x = year, y = y_value, group = factor(year_of_
    treatment),
                color = factor(year_of_treatment)),
            size = 2) +
  labs(x = "Year", y = "Value", color = "Treatment group   ") +
  theme(legend.position = 'bottom')

plot1
#ggsave(filname="sim_twfe.eps",  plot1)



##############################
# Monte Carlo Simulation
##############################


N_sim = 1000
coef_results = matrix(0, nrow = N_sim, ncol = 1)

for(s in 1:N_sim){
  df = DGP()
  twfe = plm(y_value~treat, data=df, model = 'within',effect='
    twoways')
  coef_results[s,1] = coef(twfe)["treat"]
}
```

```r
monte_sim = data.frame(coef = coef_results) %>% ggplot() +
  geom_histogram(aes(x = coef))
t.test(coef_results, mu=mu)

#ggsave("monte_sim.eps", monte_sim)


#############################
# Using the did package
#############################

df = DGP() %>% select(unit, year, year_of_treatment, y_value)

CS.ATT = att_gt(
  yname = "y_value",
  tname = "year",
  idname = "unit",
  gname = "year_of_treatment",
  control_group = "notyettreated",
  bstrap = FALSE,
  data = df)

event_std = aggte(CS.ATT, type = 'dynamic')

did_package_result =
  data.frame(TE = event_std$att.egt, year = -14:14,
             CI_UP = event_std$att.egt + event_std$se.egt,
             CI_LO = event_std$att.egt - event_std$se.egt) %>%
  ggplot() +
  geom_point(aes(x = year, y=TE), size=2) +
  geom_ribbon(aes(x = year, ymin = CI_LO, ymax = CI_UP), alpha=0.2,
   color = "lightgrey")+
  xlab("Relative year")+ylab("ATT(t,g) Estimated")
did_package_result

#ggsave("did_package_result.eps", did_package_result)
```