Double-Robust Two-Way-Fixed-EffectsRegression For Panel Data

19331053 纪传宇

2023 年 8 月 2 日

1 Motivation

1.1 Shortage of based TWFE models

We study estimation of causal effects of a binary treatment in a panel data setting with a large number of units, a modest (fixed) number of time periods, and general treatment patterns.

(1)Recent research shows that regression estimators for average treatment effects based on TWFE models might have undesirable properties, in particular, negative weights for unit-time specific treatment effects. These concerns are particularly salient in settings with heterogeneity in treatment effects and general assignment patterns (e.g., De Chaisemartin and d'Haultfoeuille [2020], Goodman-Bacon [2018], Abraham and Sun [2018], Callaway and Sant'Anna [2018], Borusyak and Jaravel [2017]).

(2)The intuition for the failure of the standard inverse propensity score weighting is that the TWFE regression model does not correspond to a consistently estimable conditional expectation because it includes unit fixed effects.

(3)We cannot expect the TWFE model or the assignmentmodel to be fully correct.

Following much of the applied work we focus on least squares estimators with two-way fixed effects (TWFE). We augment this specification with unit-specific weights, leading to the following estimator:

$$\hat{\tau}(\gamma) = \underset{\tau,\alpha_i,\lambda_t,\beta}{\arg\min} \sum_{it} \left(Y_{it} - \alpha_i - \lambda_t - \beta^\top X_{it} - \tau W_{it}\right)^2 \gamma_i \tag{1}$$

Here Y_{it} is the outcome variable of interest, W_{it} is a binary treatment, and X_{it} are observed exogenous characteristics. Unit-specific weighs γ_i are constructed using both attributes and realized assignment paths $\boldsymbol{W}_i = (W_{i1}, \ldots, W_{iT})$, but free of dependence on the outcomes.

We use this assignment model to construct the weights $\gamma^\star$ that guarantee that $\hat{\tau}(\gamma^\star)$ converges to the average (equally over units and periods) treatment effect even if the TWFE regression model is misspecified.

In general, we characterize the limiting behavior of $\hat{\tau}(\gamma)$ for a large class of weighting functions and provide an analytic correspondence between the choice of weights and the resulting causal estimand.

In applications, this model provides a parsimonious approximation for the baseline outcomes, allowing researchers to capture unobserved confounders and to improve the efficiency of the resulting estimator by reducing noise. At the same time, Our results show that some of the concerns raised in this literature regarding negative weights disappear once we properly reweight the observations.

Our estimator is robust toarbitrary violations of parallel trends assumptions, as long as the assignment model is correctlyspecified.

We show that our estimator is more robust than the conventional two-way estimator: it remains consistent if either the assignment mechanism or the two-way regression model is correctly specified and performs better than the twoway-fixed-effect estimator if both are locally misspecified.

2 Notation

We define the unit and time-specific treatment effect as:

$$\tau_{it} \triangleq Y_{it}(1) - Y_{it}(0). \tag{2}$$

For each time period t , we define the time-specific ATE as:

$$\tau_t \triangleq \frac{1}{n}\sum_{i=1}^{n} \tau_{it}, \tag{3}$$

and consider a broad class of weighted average of time-specific ATE:

$$\tau^*(\xi) \triangleq \sum_{t=1}^{T} \xi_t \tau_t \tag{4}$$

for some user-specified deterministic weights $\xi = (\xi_1, \ldots, \xi_T)^\top$ such that

$$\sum_{t=1}^{T} \xi_t = 1, \quad \xi_t \geq 0. \tag{5}$$

We refer to $\tau^*(\xi)$ as a doubly average treatment effect (DATE). For example, the weights $\xi_t = 1/T$ yield the usual ATE over units and time periods. In the difference-in-differences setting with two time periods, $\xi_t = 1_{t=2}$.

For each unit i and a possible assignment path $\boldsymbol{W}_i$ we define the generalized propensity score (Imbens [2000], Athey and Imbens [2018], Bojinov et al. [2020a,b]) - the marginal probability of such path:

$$\boldsymbol{\pi}_i(\boldsymbol{w}) = \mathbb{P}\left[\boldsymbol{W}_i = \boldsymbol{w}\right], \quad \forall \boldsymbol{w} \in \{0,1\}^T \tag{6}$$

Given our focus on design-based inference we treat $\boldsymbol{\pi}_i$ as known objects. These functions are unit-specific thus allowing for general experimental designs, for example, stratification based on observed unit characteristics. We impose minimal overlap restrictions on each $\boldsymbol{\pi}_i$

3 Reshaped IPW estimator

We consider a class of weighted TWFE regression estimators. We refer to them as reshaped inverse propensity weighted (RIPW) estimators, and formally define them as follows:

$$\hat{\tau}(\mathbf{\Pi}) \triangleq \underset{\tau,\mu,\sum_i \alpha_i=\sum_t \lambda_t=0}{\arg\min} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(Y_{it}-\mu-\alpha_i-\lambda_t-W_{it}\tau\right)^2 \frac{\mathbf{\Pi}\left(\boldsymbol{W}_i\right)}{\boldsymbol{\pi}_i\left(\boldsymbol{W}_i\right)}, \tag{7}$$

where $\mathbf{\Pi}(\boldsymbol{w})$ is a density function on $\{0,1\}^T$, i.e.,

$$\sum_{\boldsymbol{w}\in\{0,1\}^T} \Pi(\boldsymbol{w})=1 \tag{8}$$

We refer to the distribution Π as a reshaped distribution, and the weight $\Pi\left(\boldsymbol{W}_i\right)/\boldsymbol{\pi}_i\left(\boldsymbol{W}_i\right)$ as a RIP weight. To ensure that the RIPW estimator is well-defined, we require Π to be absolutely continuous with respect to each $\boldsymbol{\pi}_i$, i.e.

$$\mathbf{\Pi}(\boldsymbol{w})=0 \text{ if } \boldsymbol{\pi}_i(\boldsymbol{w})=0 \text{ and } \boldsymbol{W}_i=\boldsymbol{w} \text{ for some } i\in[n] \tag{9}$$

4 Double-Robust Inference

4.1 Double Robustness

Define two non-nested models:

(1) **The assignment model** is characterized by the generalized propensity score, defined as

$$\boldsymbol{\pi}_i(\boldsymbol{w})=\mathbb{P}\left[\boldsymbol{W}_i=\boldsymbol{w} \mid \boldsymbol{X}_i,\boldsymbol{U}_i\right], \quad \forall \boldsymbol{w}\in\{0,1\}^T \tag{10}$$

With unobserved $\boldsymbol{U}_i$, it is generally impossible to get an accurate estimate of $\boldsymbol{\pi}_i$. However, it can be constructed under additional structural assumptions. For instance, suppose that $U_{it}\equiv U_i$ is a time-invariant confounder and $(W_{i1},\ldots,W_{iT})$ are independent with

$$\operatorname{logit}\left(\mathbb{P}\left(W_{it}=1 \mid X_{it},\boldsymbol{U}_i\right)\right)=X_{it}^{\top}\beta+\gamma\left(\boldsymbol{U}_i\right) \tag{11}$$

The term $\gamma\left(\boldsymbol{U}_i\right)$ is essentially a fixed effect and cannot be estimated consistently when T=O(1) since there are. It is easy to demonstrate that $\boldsymbol{W}_i \perp\!\!\!\perp U_i \mid \boldsymbol{X}_i,\bar{W}_i$ and

$$\boldsymbol{\pi}_i(\boldsymbol{w}) \propto \frac{\exp\left\{\sum_{t=1}^{T} w_t X_{it}^{\top}\beta\right\}}{\sum_{\boldsymbol{j}\in\{0,1\}^T:\bar{\boldsymbol{j}}=\overline{\boldsymbol{w}}} \exp\left\{\sum_{t=1}^{T} j_t X_{it}^{\top}\beta\right\}} \cdot I\left\{\overline{\boldsymbol{w}}=\bar{W}_i\right\}. \tag{12}$$

The coefficient vector β can be consistently estimated via the conditional logistic regression

(2) **The outcome model** considered in this paper is a TWFE model. Specifically, the outcome model assumes that

$$\mathbb{E}\left[Y_{it}(w) \mid \boldsymbol{X}_i,\boldsymbol{U}_i\right]=\alpha\left(\boldsymbol{U}_i\right)+\lambda_t+m\left(X_{it},U_{it}\right)+\tau^* w \tag{13}$$

In particular, this implies a constant treatment effect. When T=O(1) , the unit fixed effect $\alpha\left(\boldsymbol{U}_i\right)$ cannot be estimated consistently without further assumptions on $\alpha(\cdot)$ and $\boldsymbol{U}_i$, because there are only T samples that carry information on $\alpha\left(\boldsymbol{U}_i\right)$. Thus we cannot hope to estimate $\mathbb{E}\left[Y_{it}(w) \mid \boldsymbol{X}_i, \boldsymbol{U}_i\right]$ consistently even with infinite sample sizes.

Let m_{it} denote the doubly-centered version of $\left\{\mathbb{E}\left[Y_{it}(0) \mid \boldsymbol{X}_i, \boldsymbol{U}_i\right] : i \in [n], t \in [T]\right\}$, i.e.

$$\begin{aligned} m_{it} \triangleq & \mathbb{E}\left[Y_{it}(0) \mid \boldsymbol{X}_i, \boldsymbol{U}_i\right]-\frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[Y_{it}(0) \mid \boldsymbol{X}_i, \boldsymbol{U}_i\right]-\frac{1}{T} \sum_{t=1}^T \mathbb{E}\left[Y_{it}(0) \mid \boldsymbol{X}_i, \boldsymbol{U}_i\right] \\ & +\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \mathbb{E}\left[Y_{it}(0) \mid \boldsymbol{X}_i, \boldsymbol{U}_i\right] \end{aligned} \tag{14}$$

When the outcome model is correct, it is easy to see that m_{it} is also the doubly-centered version of terms $\left\{m\left(X_{it}, U_{it}\right) : i \in [n], t \in [T]\right\}$. Given an estimate $\hat{\mu}_{it} of \mathbb{E}\left[Y_{it}(0) \mid \boldsymbol{X}_i, \boldsymbol{U}_i\right]$,

4.2 RIPW estimators

Given an estimate $\hat{\mu}_{it}$ for $\mathbb{E}\left[Y_{it}(0) \mid \boldsymbol{X}_i, \boldsymbol{U}_i\right]$ and $\hat{\boldsymbol{\pi}}_i$ for $\boldsymbol{\pi}_i$, we consider the following RIPW estimator

$$\hat{\tau}(\boldsymbol{\Pi}) \triangleq \underset{\tau, \mu, \sum_i \alpha_i=\sum_t \gamma_t=0}{\arg\min} \sum_{i=1}^n \sum_{t=1}^T\left(\left(Y_{it}^{\text{obs}}-\hat{m}_{it}\right)-\mu-\alpha_i-\gamma_t-W_{it} \tau\right)^2 \frac{\boldsymbol{\Pi}\left(\boldsymbol{W}_i\right)}{\hat{\boldsymbol{\pi}}_i\left(\boldsymbol{W}_i\right)} . \tag{15}$$

This is more general than the following weighted TWFE regression estimator with covariates

$$\hat{\tau} \triangleq \underset{\tau, \mu, \beta, \sum_i \alpha_i=\sum_t \gamma_t=0}{\arg\min} \sum_{i=1}^n \sum_{t=1}^T\left(Y_{it}^{\text{obs}}-\mu-\alpha_i-\gamma_t-W_{it} \tau-X_{it}^{\top} \beta\right)^2 \frac{\boldsymbol{\Pi}\left(\boldsymbol{W}_i\right)}{\hat{\boldsymbol{\pi}}_i\left(\boldsymbol{W}_i\right)}, \tag{16}$$

which is a special case of (15) with $\hat{m}_{it}=X_{it}^{\top} \hat{\beta}$. The two-stage estimator (15) is more flexible since it does not require $\hat{m}_{it}$ to be estimated from the same weighted regression for DATE.

5 Solutions of the DATE equation

Analogue in the panel data:

$$\hat{\tau}_{\text{IPW}} \triangleq \underset{\tau}{\arg\min} \sum_{i=1}^n \sum_{t=1}^T \underbrace{\left(Y_{it}-\alpha_i-\lambda_t-W_{it} \tau\right)^2}_{\text{TWFE objective}} \underbrace{\frac{1}{\boldsymbol{\pi}_i\left(\mathbf{W}_i\right)}}_{\text{generalized propensity score}} \xrightarrow{p} ? \tag{17}$$

5.1 Example

Transient treatments

$$\begin{aligned} & \mathbf{W}_i \in\{(0,0,0),(0,0,1),(0,1,0),(1,0,0)\} \\ & \hat{\tau}_{\text{IPW}} \xrightarrow{p} \tfrac{1}{3} \tau_1+\tfrac{1}{3} \tau_2+\tfrac{1}{3} \tau_3=\tau_{\text{eq}} \end{aligned} \tag{18}$$

Staggered rollouts

$$\begin{aligned} & \mathbf{W}_i \in\{(0,0,0),(0,0,1),(0,1,1),(1,1,1)\} \\ & \hat{\tau}_{\mathrm{IPW}} \xrightarrow{p} 0.3 \tau_1+0.4 \tau_2+0.3 \tau_3 \end{aligned} \tag{19}$$

Theorem (Arkhangelsky, Imbens, L., and Luo '21) Under regularity conditions (overlap, limited dependence, bounded moments), as $n \rightarrow \infty$,

$$\hat{\tau}_{\mathrm{IpW}} \xrightarrow{p} \sum_{t=1}^T \xi_t \tau_t \tag{20}$$

where $\mathbb{S}=\bigcup_i \operatorname{Supp}\left(\mathbf{W}_i\right)$ and

$$\xi_t \propto \eta_t\left(1-\eta_t\right), \quad \text { where } \eta_t=\frac{\left|w \in \mathbb{S}: w_t=1\right|}{|\mathbb{S}|} \tag{21}$$

5.2 The case of two periods

When there are two periods, the DATE equation only involves four variables $\boldsymbol{\Pi}(0,0), \boldsymbol{\Pi}(0,1), \Pi(1,0), \Pi(1,1)$. Through some tedious algebra presented in Appendix B.1, we can show that the DATE equation can be simplified into the following equation:

$$\{\boldsymbol{\Pi}(1,1)-\boldsymbol{\Pi}(0,0)\}\{\boldsymbol{\Pi}(1,0)-\boldsymbol{\Pi}(0,1)\}=\left(\xi_1-\xi_2\right)\left\{(\boldsymbol{\Pi}(1,0)-\boldsymbol{\Pi}(0,1))^2-(\boldsymbol{\Pi}(1,0)+\boldsymbol{\Pi}(0,1))\right\} . \tag{22}$$

In the setting of difference-in-difference (DiD), (0,0) and (0,1) are the only two possible treatment assignments. As a result, we should set the support of the reshaped distribution to be $\mathbb{S}^*=\{(0,0),(0,1)\}$. Then (22) reduces to

$$\boldsymbol{\Pi}(0,0) \boldsymbol{\Pi}(0,1)=\left(\xi_1-\xi_2\right)\left(\boldsymbol{\Pi}(0,1)^2-\boldsymbol{\Pi}(0,1)\right)=\left(\xi_2-\xi_1\right) \boldsymbol{\Pi}(0,0) \boldsymbol{\Pi}(0,1) . \tag{23}$$

It has a solution only when $\xi_2-\xi_1=1, i.e. \left(\xi_1, \xi_2\right)=(0,1)$ and hence $\tau^*(\xi)=\tau_2$, in which case any reshaped distribution \$$Pi$ with $\Pi(0,0), \Pi(0,1)>0$ is a solution. The RIPW estimator with any Π with $\Pi(0,0), \boldsymbol{\Pi}(0,1)>0$ and $\Pi(0,0)+\Pi(0,1)=1$ yields a double-robust DiD estimator.So if the distribution is uniform, then ξ_1 is 0 and ξ_2 is 1 which is as usual is IPW method.

5.3 Staggered adoption with multiple periods

The RIPW estimator with a uniform reshaped distribution is inconsistent, the non-uniform distribution (1 / 3,1 / 6,1 / 6,1 / 3) , namely the midpoint of the solution set, induces a consistent RIPW estimator. For general T , it is easy to see that the midpoint is

$$\boldsymbol{\Pi}\left(\boldsymbol{w}_{(T)}\right)=\boldsymbol{\Pi}\left(\boldsymbol{w}_{(0)}\right)=\frac{T+1}{4 T}, \boldsymbol{\Pi}\left(\boldsymbol{w}_{(j)}\right)=\frac{1}{2 T}, j=1, \ldots, T-1 \tag{24}$$

This distribution uniformly assigns probabilities on the subset $\left\{\boldsymbol{w}_{(1)}, \ldots, \boldsymbol{w}_{(T-1)}\right\}$ while puts a large mass on $\left\{\boldsymbol{w}_{(0)}, \boldsymbol{w}_{(T)}\right\}$. Intuitively, the asymmetry is driven by the special roles of $\boldsymbol{w}_{(0)}$ and $\boldsymbol{w}_{(T)}$: the former provides the only control group for period T while the latter provides the only treated group for period 1 .

6 Example

Reanalysis of Bachhuber et al. (2014) on medical cannabis law As more states pass the medical cannabis law, there has been debate on whether legal medical marijuana is associated with an increase or decrease in opioid overdose mortality.

Following Bachhuber et al. [2014] and Shover et al. [2019], we use the logarithm of ageadjusted opioid overdose death rate per 100,000 population as the outcome, and include four time-varying covariates: annual state unemployment rate and presence of the following: prescription drug monitoring program, pain management clinic oversight laws, and law requiring or allowing pharmacists to request patient identification.

For the RIPW estimator, we fit a standard TWFE regression to derive an estimate of the outcome model, i.e., $\hat{m}_{it} = X_{it}^{\top}\hat{\beta}$. This step guarantees that the resulting RIPW estimator is acceptable if the analyses of Bachhuber et al. [2014] and Shover et al. [2019] are because all estimate the same outcome model.

```r
## Estimate mhat from a two-way fixed effect regression
## without cross-fitting
twfe_mhat_onefold <- function(Y, tr, X,
testX = NULL,
timevarying = FALSE,
interact = FALSE,
joint = FALSE,
unitfe = TRUE,
timefe = TRUE,
usefe = FALSE){
       n <- nrow(Y)
       T <- ncol(Y)
       obj <- twfe_std_input(Y, tr, X, timevarying, interact, joint,
           unitfe, timefe)
       if (is.null(testX)){
               overlap <- TRUE
               testX <- obj$X
       } else {
               overlap <- FALSE
               if (usefe){
                       stop("Fixed effects cannot be estimated when X
                           is not the same as testX")
               }
       }
       if (!is.null(X)){
               n_test <- floor(nrow(testX) / T)
       }
```

```
        if (!joint){
            if (usefe){
                stop("Fixed effects cannot be estimated for
                    separate fitting")
            }
            if (is.null(X)){
                mhat1 <- matrix(0, n, T)
                mhat0 <- matrix(0, n, T)
            } else {
                fit1 <- lfe::felm(as.formula(obj$formula), data
                     = obj$data, subset = (tr == 1))
                fit0 <- lfe::felm(as.formula(obj$formula), data
                     = obj$data, subset = (tr == 0))
                cf1 <- as.numeric(fit1$coef)
                cf1[is.nan(cf1)] <- 0
                cf0 <- as.numeric(fit0$coef)
                cf0[is.nan(cf0)] <- 0
                mhat1 <- matrix(testX %*% cf1, n_test, T)
                mhat0 <- matrix(testX %*% cf0, n_test, T)
            }
        } else {
            fit <- lfe::felm(as.formula(obj$formula), data = obj$
                data)
            cf <- as.numeric(head(fit$coef, -1))
            tau <- as.numeric(tail(fit$coef, 1))
            fe <- lfe::getfe(fit)
            timefe <- fe$effect[(1:T) + n]
            if (!overlap || !usefe){
                if (!is.null(X)){
                    mu <- mean(fitted.values(fit) - tau *
                        as.numeric(tr) - obj$X %*% cf - rep(
                        timefe, each = n))
                    mhat0 <- testX %*% cf + mu + rep(timefe
                        , each = n_test)
                    mhat0 <- matrix(mhat0, n_test)
                } else {
                    mu <- mean(fitted.values(fit) - tau *
                        as.numeric(tr) - rep(timefe, each =
                        n))
                    mhat0 <- rep(timefe, each = n) + mu
                    mhat0 <- matrix(mhat0, n)
                }
            } else {
```

```
                        mhat0 <- fitted.values(fit)
                        if (is.null(X)){
                                mhat0 <- matrix(mhat0, n)
                        } else {
                                mhat0 <- matrix(mhat0, n_test)
                        }
                }
                mhat1 <- mhat0 + tau
        }
        mhat <- list(mhat1 = mhat1, mhat0 = mhat0)
        return(mhat)
}
```

On top of that, we fit a Cox proportional hazard model [Cox, 1972, Kalbfleisch and Prentice, 2011] with the same set of covariates to model the rightcensored adoption time. Specifically, letting T_i be the year in which the state i passes the medical cannabis law, a Cox proportional hazard model with time-varying covariates X_{it} assumes that

$$h_i\left(t \mid X_{it}\right) = h_0(t) \exp\left\{X_{it}^{\top}\beta\right\} \tag{25}$$

where $h_i(t \mid \cdot)$ denotes the hazard function for state i , and $h_0(t)$ denotea a nonparametric baseline hazard function.

```
survdata <- list()
for (i in 1:length(states)){
        tmpdata <- data %>%
        filter(state == states[i]) %>%
        select(treat, time,
        unemployment, vote,
        rxdmp_original,
        rxid_original,
        pmlaw_original,
        year)
        inds <- which(tmpdata$treat > 0)
        if (length(inds) > 0){
                ind <- min(inds)
                tmpdata <- tmpdata[1:ind, ] %>%
                mutate(time1 = time - 1,
                time2 = time) %>%
                select(-time) %>%
                mutate(status = 0,
                id = i)
                tmpdata$status[length(tmpdata$status)] <- 1
        } else {
                tmpdata <- tmpdata %>%
```

```
                        mutate(time1 = time - 1,
                        time2 = time) %>%
                        select(-time) %>%
                        mutate(status = 0,
                        id = i)
                }
                survdata[[i]] <- tmpdata
        }
        survdata <- do.call(rbind, survdata) %>%
        select(unemployment, vote, rxdmp_original, rxid_original, pmlaw_
            original, year, time1, time2, status, id)
```

The proportional hazard assumption required by the Cox model is sometimes questioned. In this study, we utilize the conventional statistical tests that rely on Schoenfeld residuals [Schoenfeld, 1980] to serve as a diagnostic tool for checking the appropriateness of the Cox model's specification.

```
     ## test for proportional hazard assumption
    fit <- coxph(Surv(time1, time2, status) ~ unemployment + rxdmp_original +
        rxid_original + pmlaw_original, data = survdata)
    test <- cox.zph(fit)
    schoenfeld_pval <- as.numeric(tail(test$table[, 3], 1))
```

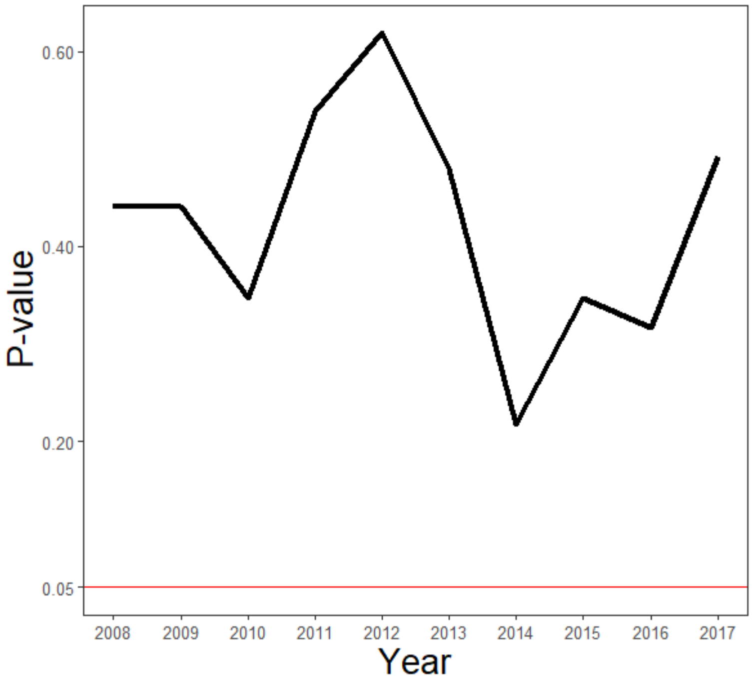

图 1: Diagnostics for the Cox proportional hazard model on adoption times;

This figure shows that the p-values yielded by the Schoenfeld' s test for each Tend without data splitting. Clearly, none of them show evidence against the proportional hazard assumption.

The estimates $\hat{h}_0$ and $\hat{\beta}$ yield an estimate $\hat{F}_i(t)$ of the survival function $\mathbb{P}\left(T_i \geq t\right)$ for state i , differencing which yields an estimate generalized propensity score

$$\hat{\boldsymbol{\pi}}_i\left(\boldsymbol{W}_i\right)=\begin{cases}\hat{F}_i\left(T_i\right)-\hat{F}_i\left(T_i+1\right) & (\text{ State } i \text{ passed the law before 2017}) \\ 1-\hat{F}_i(2017) & (\text{otherwise})\end{cases}. \tag{26}$$

```
## Estimate generalized propensity scores with cross-fitting
gps_est <- rep(NA, nstates)
adopt_time <- rep(NA, nstates)
adopt <- rep(NA, nstates)
for (k in 1:nfolds){
        ids <- foldid[[k]]
        train <- filter(survdata, !(id %in% ids))
        test <- filter(survdata, id %in% ids)
        status_df <- test %>% group_by(id) %>%
        summarize(status = max(status))
        adopt[ids] <- status_df$status
        fit <- coxph(Surv(time1, time2, status) ~ unemployment + rxdmp_
            original + rxid_original + pmlaw_original, data = train)
        obj <- summary(survfit(fit, newdata = test, id = id))
        for (index in ids){
                status <- status_df$status[status_df$id == index]
                if (is.null(obj$strata)){
                        survtime <- obj$time
                        survprob <- obj$surv
                } else {
                        survtime <- obj$time[obj$strata == index]
                        survprob <- obj$surv[obj$strata == index]
                }
                if (status == 1){
                        if (length(survprob) == 1){
                                gps_est[index] <- 1 - survprob
                        } else {
                                gps_est[index] <- -diff(tail(survprob,
                                    2))
                        }
                } else {
                        gps_est[index] <- tail(survprob, 1)
                }
                if (status){
                        adopt_time[index] <- max(survtime)
                }
        }
}
```

```r
## Reshaped distribution
T <- max(data$year) - min(data$year) + 1
adopt_time[is.na(adopt_time)] <- T + 1
adopt_time <- T + 1 - adopt_time

Pi <- rep(0, T + 1)
Pi[2:T] <- 1 / 2 / T
Pi[c(1, T+1)] <- (T + 1) / 4 / T

EPi <- head(cumsum(rev(Pi)), -1)
Theta <- Pi[adopt_time + 1] / gps_est
```

The reshaped distribution is chosen as the midpoint solution . Finally, we apply the standard 10 -fold cross-fitting to derive the estimates of the outcome and treatment models.

```r
## Reshaped distribution
T <- max(data$year) - min(data$year) + 1
adopt_time[is.na(adopt_time)] <- T + 1
adopt_time <- T + 1 - adopt_time

Pi <- rep(0, T + 1)
Pi[2:T] <- 1 / 2 / T
Pi[c(1, T+1)] <- (T + 1) / 4 / T

EPi <- head(cumsum(rev(Pi)), -1)
Theta <- Pi[adopt_time + 1] / gps_est

## Generate the outcomes and treatment assignments
Y <- data %>% select(state, year, mortality) %>%
spread(year, mortality) %>%
select(-state) %>%
as.matrix
tr <- data %>% select(state, year, treat) %>%
spread(year, treat) %>%
select(-state) %>%
as.matrix

X <- data %>%
select(treat, unemployment,
rxdmp_original, rxid_original, pmlaw_original) %>%
as.matrix
X <- lapply(1:T, function(j){
   X[(j-1)*nstates + 1:nstates, ]
```

```
})
X <- Reduce(cbind, X)
muhat <- twfe_mhat(Y, tr, X,
timevarying = TRUE,
joint = TRUE,
foldid = foldid)
obj_ripw <- ripw(Y, tr, muhat = muhat$mhat0, Theta = Theta)
```

We calculate the RIPW estimates of equally-weighted DATE, the unweighted TWFE regression estimates for $T_{\text{end}} \in \{2008, 2009, \ldots, 2017\}$ and the 95 % pointwise confidence intervals; here, the cluster-robust standard error is used for the unweighted estimator.

```
res_unw <- res %>%
select(year, contains("unw")) %>%
mutate(lb = unw_tauhat - 1.96 * unw_se,
ub = unw_tauhat + 1.96 * unw_se) %>%
rename(tauhat = unw_tauhat) %>%
mutate(estimator = "UNW") %>%
select(estimator, year, tauhat, lb, ub)
res_ripw <- res %>%
select(year, contains("ripw")) %>%
mutate(lb = ripw_tauhat - 1.96 * ripw_se,
ub = ripw_tauhat + 1.96 * ripw_se) %>%
rename(tauhat = ripw_tauhat) %>%
mutate(estimator = "RIPW") %>%
select(estimator, year, tauhat, lb, ub)
plot <- rbind(res_unw, res_ripw) %>%
ggplot(aes(x = year, y = tauhat, color = estimator, linetype = estimator))
   +
geom_line(size = 1.5) +
geom_hline(yintercept = 0, col = "black") +
geom_ribbon(aes(ymin = lb, ymax = ub, color = estimator, fill = estimator,
   linetype = estimator), alpha = 0.25) +
scale_x_continuous(breaks = 2008:2017,
expand = c(0, 0, 0, 0)) +
scale_color_discrete(name = "Estimator") +
scale_linetype_discrete(name = "Estimator") +
scale_fill_discrete(name = "Estimator") +
xlab("Year") + ylab("Estimate") +
theme_bw() +
theme(panel.grid = element_blank(),
text = element_text(size = 20),
axis.text = element_text(size = 10),
legend.key.width = unit(1.25, "cm"))
```

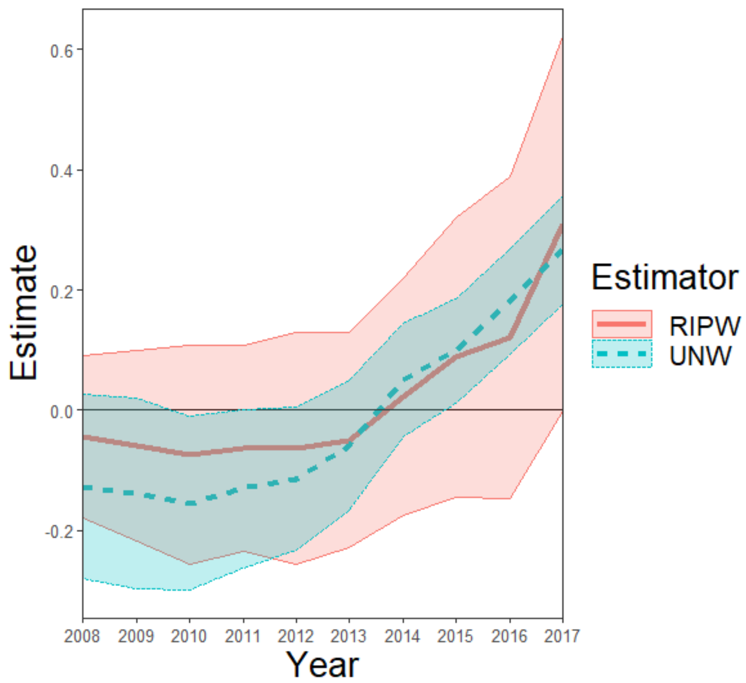

图 2: Diagnostics for the Cox proportional hazard model on adoption times;