

A Vignette for LATE Estimation

Baoluo Sun¹ & Zhiqiang Tan²

October 2020

1 Introduction

Identification of the average treatment effect (ATE) relies on treatment unconfoundedness given the measured covariates. In the presence of unmeasured confounding, an alternative approach makes use of an exogenous experimental called an instrumental variable (IV) to identify under certain conditions the local average treatment effect (LATE), defined as the average treatment effect among individuals whose treatment status would be manipulated through the change of the IV (Angrist et al. 1996). The R package `RCAL` (version 2.0) includes three functions for estimating LATE, similarly to estimation of ATE discussed in a separate vignette:

- `late.nreg`: inference using non-regularized calibrated estimation,
- `late.regu.cv`: inference using regularized calibrated estimation based on cross validation,
- `late.regu.path`: inference using regularized calibrated estimation along a regularization path.

Both the treatment and IV are assumed to be binary. Extensions to multi-valued treatments and IVs will be incorporated in later versions.

2 An example

We illustrate the use of the package on a simulated dataset as in Sun and Tan (2020), Section 4. The dataset, `simu.iv.data`, is included as part of the package.

¹Department of Statistics and Applied Probability, National University of Singapore, Singapore, SG 117546. Email: stasb@nus.edu.sg.

²Department of Statistics, Rutgers University. Address: 110 Frelinghuysen Road, Piscataway, NJ 08854. E-mail: ztan@stat.rutgers.edu.

```
> library(RCAL)
> data(simu.iv.data)
```

The following shows the first 10 rows and the first 6 columns of the dataset, which is of size 800×203 .

```
> simu.iv.data[1:10, 1:6]
```

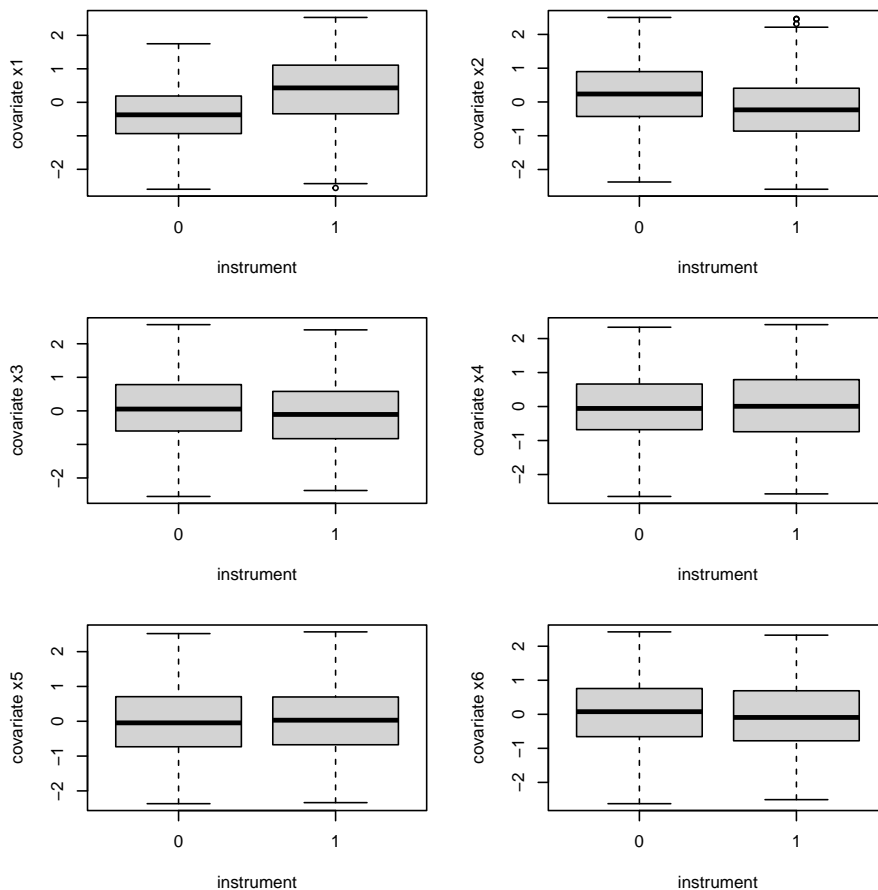
```
      y tr iv
[1,] -4.2383798  1  1 -0.2873990  1.61831773 -0.5806894
[2,] -0.6003803  0  1 -0.4189257 -0.35481518 -1.4303633
[3,]  6.9089746  0  0 -0.8966501  2.23983788 -0.3732223
[4,]  7.0108296  1  1  2.1739207 -2.31342276  0.7050739
[5,] -2.3579281  1  0  1.0981814 -1.52994363 -1.1181420
[6,] 12.4793401  0  0  0.2367108  0.87265233  0.9561064
[7,]  8.8237297  1  0 -0.9480914  2.08710627 -0.5260312
[8,] -3.1041262  0  1  1.7899789  0.03542566 -1.7308174
[9,]  5.8488491  0  0  0.1031749 -0.24369528 -0.2866433
[10,] 3.0499787  0  0  0.8340469 -0.73645791 -0.3628298
```

The first column represents an observed outcome y , the second column represents a binary treatment tr , the third column represents a binary IV iv , and the remaining 200 columns represent covariates.

```
> n <- dim(simu.iv.data)[1]
> p <- 100 # include the first 100 covariates due to CRAN time constraint
> y <- simu.iv.data[,1]
> tr <- simu.iv.data[,2]
> iv <- simu.iv.data[,3]
> x <- simu.iv.data[,3+1:p]
> x <- scale(x)
```

To examine the data, Figure 1 shows the boxplots of the first 6 covariates in the $iv=0$ and $iv=1$ groups, Figure 2 shows the boxplots of the first 6 covariates in

Figure 1: Boxplots of covariates in the $iv=0$ and $iv=1$ groups.

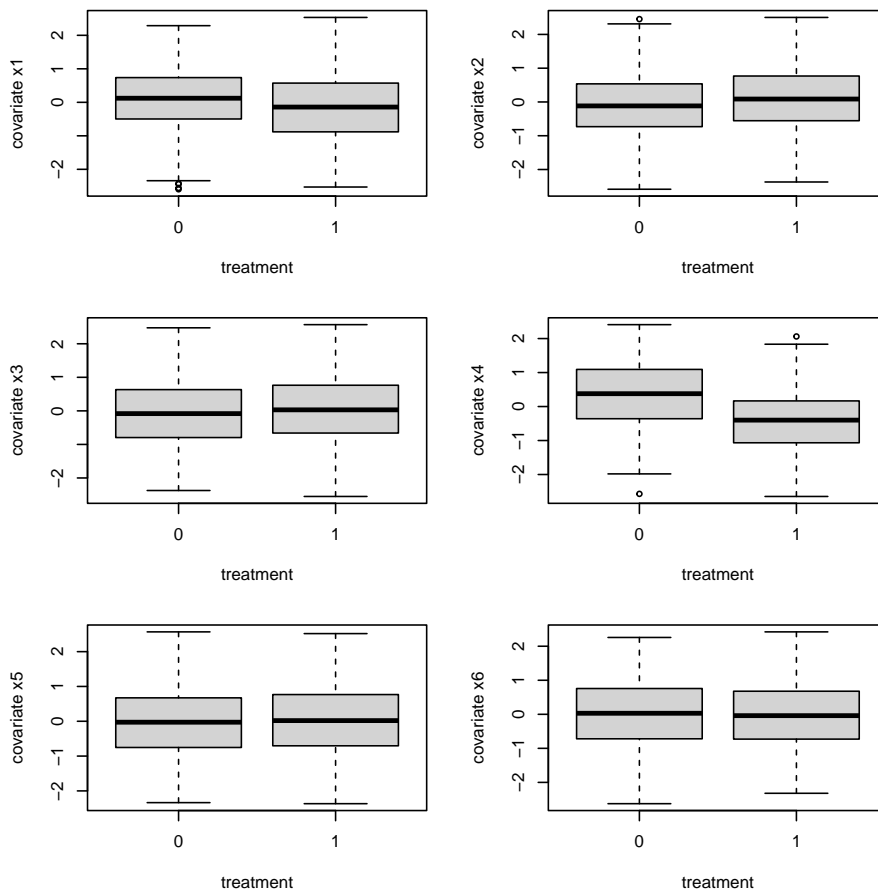


the untreated and treated groups, and Figure 3 shows the scatterplots of observed outcomes and the first 6 covariates in the treated group.

The dataset is generated from a parametric latent index model conditionally on covariates as described in `help(simu.iv.data)`. The true value of LATE is 1 by design. Vytlacil (2002) and Tan (2006) showed that the structural assumptions for identification of LATE are equivalent to the assumptions of a nonparametric latent index model. Ignoring the covariates, the Wald estimator yields an estimate 1.48 with standard error 0.766.

```
> pi    <- mean(iv)
> del   <- iv/pi - (1-iv)/(1-pi);
> del.y <- mean(del*y);
```

Figure 2: Boxplots of covariates in the untreated and treated groups.



```

> del.tr<- mean(del*tr);
> (w<- del.y/del.tr) # point estimate

[1] 1.478608

> g <- mean((-iv/pi^2-(1-iv)/(1-pi)^2)*(y-w*tr))
> sqrt(mean((del*(y-w*tr)-g*(pi-iv))^2)/del.tr^2/n) # standard error

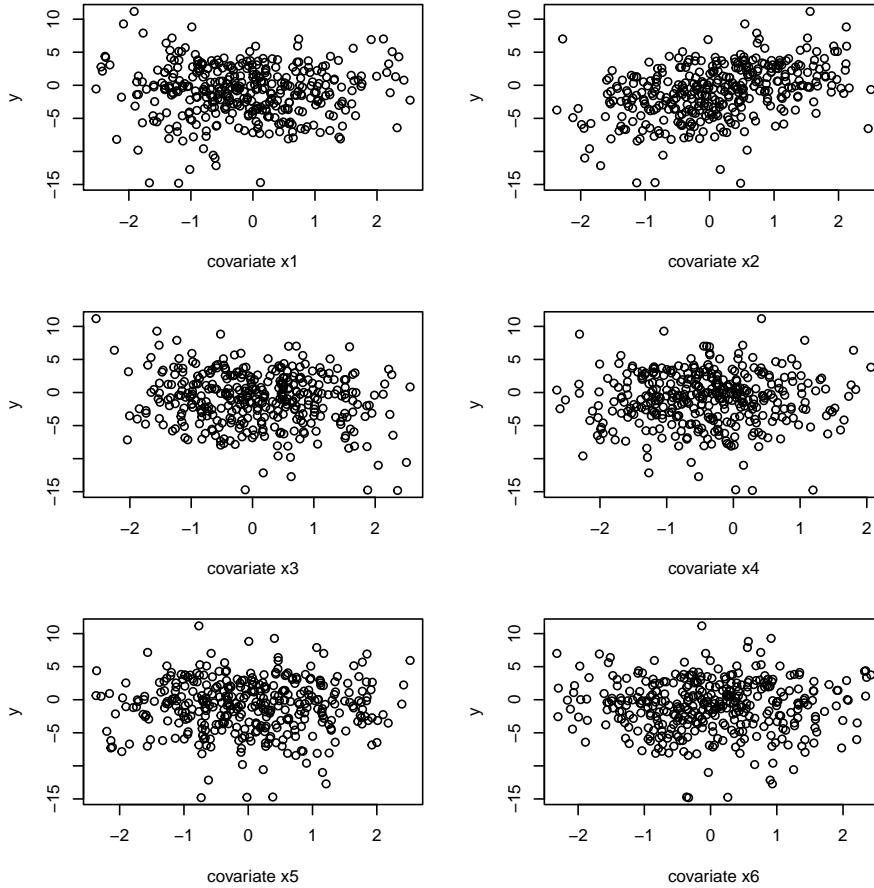
[1] 0.7659565

```

2.1 Regularized estimation of LATE

We use the potential outcomes notation (Neyman 1990; Rubin 1974) to define quantities of causal interest. For $z, d \in \{0, 1\}$ and each individual i , the potential treatment

Figure 3: Scatterplots of observed outcomes and covariates in the treated group.



status $D_i(z)$ is the observed treatment D_i if the instrument Z_i takes on value z , and the potential outcome $Y_i(d, z)$ is the observed outcome Y_i if the treatment and instrument D_i and Z_i take on values d and z respectively. The covariates X_i are observed on all individuals in the sample. The two basic IV assumptions are instrument unconfoundedness and exclusion restriction $Y_i(d, 1) = Y_i(d, 0)$, henceforth denoted as $Y_i(d)$, for $d \in \{0, 1\}$. Then under an additional monotonicity assumption and other technical conditions, the population LATE, $\theta = E\{Y(1) - Y(0) | D(1) > D(0)\}$, is identified from observed data nonparametrically (Angrist et al. 1996).

Tan (2006) derived identification formulae for the individual expectations $\theta_d = E\{Y(d) | D(1) > D(0)\}$ for $d \in \{0, 1\}$, which can be informative in addition to the

difference $\theta = \theta_1 - \theta_0$ in applications. The formula for θ_1 is

$$\theta_1 = \frac{E\{D(1)Y(1)\} - E\{D(0)Y(1)\}}{E\{D(1)\} - E\{D(0)\}}, \quad (1)$$

which is a ratio of two differences, depending on potential outcomes and treatments. By instrument unconfoundedness, each expectation in the numerator and denominator of (1) can be estimated through augmented IPW estimation based on the observed data (O_1, \dots, O_n) where $O_i = (Y_i, D_i, Z_i, X_i)$, in parallel to related estimation methods under the assumption of treatment unconfoundedness (Robins et al. 1994; Tan 2007). For this purpose, several regression functions are involved. On one hand, the covariates X_i can be associated with the IV Z_i , which is captured by the conditional probability $\pi(X_i) = P(Z_i = 1|X_i)$ called the instrument propensity score (IPS) (Tan 2006). On the other hand, the covariates X_i can also be associated with the treatment and outcome variables within each level of the instrument. Let $m_z(X_i) = P(D_i = 1|Z_i = z, X_i)$ and $m_{1z}(X_i) = E(Y_i|D_i = 1, Z = z, X_i)$, called the treatment regression function and the outcome regression function in the treated respectively for $z \in \{0, 1\}$.

Let $\hat{\pi}_1$ and $\hat{\pi}_0$ be two possibly different versions of fitted values for the IPS, and let \hat{m}_z and \hat{m}_{1z} be the fitted treatment and outcome regression functions respectively for $z \in \{0, 1\}$. The expectations $E\{D(1)\}$ and $E\{D(0)\}$ in (1) can be estimated by $n^{-1} \sum_{i=1}^n \{\varphi_{D_1}(O_i; \hat{\pi}_1, \hat{m}_1)\}$ and $n^{-1} \sum_{i=1}^n \{\varphi_{D_0}(O_i; \hat{\pi}_0, \hat{m}_0)\}$ respectively, where

$$\varphi_{D_1}(O; \hat{\pi}_1, \hat{m}_1) = \frac{Z}{\hat{\pi}_1(X)} D - \left\{ \frac{Z}{\hat{\pi}_1(X)} - 1 \right\} \hat{m}_1(X), \quad (2)$$

$$\varphi_{D_0}(O; \hat{\pi}_0, \hat{m}_0) = \frac{1-Z}{1-\hat{\pi}_0(X)} D - \left\{ \frac{1-Z}{1-\hat{\pi}_0(X)} - 1 \right\} \hat{m}_0(X). \quad (3)$$

The expectations $E\{D(1)Y(1)\}$ and $E\{D(0)Y(1)\}$ in (1) can be estimated by $n^{-1} \sum_{i=1}^n \{\varphi_{D_1 Y_{11}}(O_i; \hat{\pi}_1, \hat{m}_1, \hat{m}_{11})\}$ and $n^{-1} \sum_{i=1}^n \{\varphi_{D_0 Y_{10}}(O_i; \hat{\pi}_0, \hat{m}_0, \hat{m}_{10})\}$ respectively, where

$$\varphi_{D_1 Y_{11}}(O; \hat{\pi}_1, \hat{m}_1, \hat{m}_{11}) = \frac{Z}{\hat{\pi}_1(X)} DY - \left\{ \frac{Z}{\hat{\pi}_1(X)} - 1 \right\} \hat{m}_1(X) \hat{m}_{11}(X), \quad (4)$$

$$\varphi_{D_0 Y_{10}}(O; \hat{\pi}_0, \hat{m}_0, \hat{m}_{10}) = \frac{1-Z}{1-\hat{\pi}_0(X)} DY - \left\{ \frac{1-Z}{1-\hat{\pi}_0(X)} - 1 \right\} \hat{m}_0(X) \hat{m}_{10}(X). \quad (5)$$

By (1), the resulting augmented IPW estimator of θ_1 is

$$\hat{\theta}_1 = \frac{n^{-1} \sum_{i=1}^n \{\varphi_{D_1 Y_{11}}(O_i; \hat{\pi}_1, \hat{m}_1, \hat{m}_{11}) - \varphi_{D_0 Y_{10}}(O_i; \hat{\pi}_0, \hat{m}_0, \hat{m}_{10})\}}{n^{-1} \sum_{i=1}^n \{\varphi_{D_1}(O_i; \hat{\pi}_1, \hat{m}_1) - \varphi_{D_0}(O_i; \hat{\pi}_0, \hat{m}_0)\}}. \quad (6)$$

Estimation of θ_0 can be similarly handled but require fitted outcome regression functions in the untreated.

Two regularized methods are implemented by the function `late.regu.cv` for fitting IPS, treatment and outcome regressions. The fitted values are then substituted into (6) to estimate θ_1 and similarly to estimate θ_0 . For `ploss="cal"`, regularized calibrated estimation of the IPS and regularized weighted likelihood estimation of the treatment and outcome regression models are performed. The method leads to model-assisted inference, in which confidence intervals are valid with high-dimensional data if the IPS model is correctly specified, but the treatment and outcome regression models may be misspecified (Sun and Tan 2020). For `ploss="ml"`, regularized maximum likelihood estimation is used (Chernozhukov et al. 2018). In this case, standard errors are only shown to be valid if the IPS, treatment and outcome models are all correctly specified.

The following shows the results using the simulated dataset. Depending on `arm=0`, 1 or 2, the function `late.regu.cv` provides estimation of θ_0 , θ_1 or both (θ_0, θ_1) respectively.

```
> ## regularized calibrated estimation
> RNGversion('3.5.0')
> set.seed(0) # this affects random split of data in cross validation
> late.cv.rcal <-
+ late.regu.cv(fold=5*c(1,1,1), nrho=(1+10)*c(1,1,1), rho.seq=NULL,
+ y, tr, iv, fx=x, gx=x, hx=x, arm=2, d1=1, d2=3, ploss="cal", yloss="gaus")
> matrix(unlist(late.cv.rcal$est), ncol=2, byrow=TRUE,
+ dimnames=list(c("ipw", "or", "est", "var", "ze",
+ "late.est", "late.var", "late.ze"), c("theta1", "theta0")))

```

	theta1	theta0
ipw	0.7347164	-0.5967274

```

or      0.6121529 -0.5973384
est     0.5756415 -0.3834344
var     0.2272339  0.2764657
ze      1.2075789 -0.7292400
late.est      NA  0.9590759
late.var      NA  0.5197250
late.ze       NA  1.3303509

> ## For comparison, we use the same set of models
> ## in regularized maximum likelihood estimation
> set.seed(0)
> late.cv.rml <-
+ late.regu.cv(fold=5*c(1,1,1), nrho=(1+10)*c(1,1,1), rho.seq=NULL,
+ y, tr, iv, fx=x, gx=x, hx=x, arm=2, d1=1, d2=3, ploss="ml", yloss="gaus")
> matrix(unlist(late.cv.rml$est), ncol=2, byrow=TRUE,
+ dimnames=list(c("ipw", "or", "est", "var", "ze",
+ "late.est", "late.var", "late.ze"), c("theta1", "theta0")))

```

	theta1	theta0
ipw	0.4477703	-0.4950192
or	0.5016241	-0.3018231
est	0.4598952	-0.3222461
var	0.2205128	0.2456359
ze	0.9793592	-0.6501921
late.est	NA	0.7821413
late.var	NA	0.4733704
late.ze	NA	1.1368016

For this example, although the IPS, treatment and outcome regressions models using main effects of X_i appear adequate based on Figures 1–3, they are all (slightly) misspecified by the design of the simulated data; see `help(simu.iv.data)`. Nevertheless, regularized calibrated estimation of LATE yields an estimate 0.96 with standard

error 0.72, whereas regularized maximum likelihood estimation yields an estimate 0.78 with standard error 0.69. Both estimates are much closer to the true value 1, with smaller standard errors, than that of the Wald estimator.

2.2 Instrument propensity score for covariate calibration

Systematic differences in the covariates X_i can exist between the $Z_i = 1$ and $Z_i = 0$ groups, and direct comparisons of observed treatments and outcomes from the two groups are not appropriate. The idea of inverse probability weighting is to reweight individuals in the $Z_i = 1$ group by the inverse of IPS, in the hope that the weighted averages of covariates in the $Z_i = 1$ group are similar to or calibrated with the simple averages in the entire sample. The effect of calibration can be measured by the following differences:

$$\frac{1}{n} \sum_{i=1}^n \left\{ \frac{Z_i}{\hat{\pi}_1(X_i)} - 1 \right\} X_{ji} = \sum_{i: Z_i=1, 1 \leq i \leq n} \hat{w}_i X_{ji} - \frac{1}{n} \sum_{i=1}^n X_{ji}, \quad j = 1, \dots, p, \quad (7)$$

where $\hat{w}_i = \{n\hat{\pi}_1(X_i)\}^{-1}$ and X_{ji} denotes the j th component of X_i . If the covariates are standardized with sample means 0 and variances 1, then the above gives the standardized calibration differences as in Tan (2020a), Section 6. Similar differences can be computed for the $Z_i = 0$ group based on (7) with Z_i replaced by $1 - Z_i$ and $\hat{\pi}_1$ replaced by $1 - \hat{\pi}_0$.

The following shows the calculation of such calibration differences based on fitted IPS returned by `late.regu.cv`. The results within the groups $\{i : Z_i = 1, 1 \leq i \leq n\}$ and $\{i : Z_i = 0, 1 \leq i \leq n\}$ are plotted in Figures 4 and 5 respectively.

```
> ## Computation of standardized calibration differences
> ##                               within iv=1 group
> fips.raw <- rep(mean(iv), n)      #constant propensity scores
> fips1.cv.rcal <- late.cv.rcal$mfp[,2]
> fips1.cv.rml <- late.cv.rml$mfp[,2]
> check1.raw <- mn.ipw(x, iv, fips.raw)
> check1.cv.rcal <- mn.ipw(x, iv, fips1.cv.rcal)
```

```

> check1.cv.rml <- mn.ipw(x, iv, fips1.cv.rml)
> par(mfrow=c(2,2))
> par(mar=c(4,4,2,2))
> plot(check1.raw$est, xlim=c(0,p), ylim=c(-.3,.3),
+       xlab="", ylab="std diff", main="Raw")
> abline(h=0)
> plot(check1.cv.rml$est, xlim=c(0,p), ylim=c(-.3,.3),
+       xlab="", ylab="std diff", main="RML")
> abline(h=0)
> abline(h=max(abs(check1.cv.rml$est)) *c(-1,1), lty=2)
> nz1.rml <- which(late.cv.rml$ips[[2]]$sel.bet[,1]!=0)-1
> text(nz1.rml, -0.3, "x", cex=1)
> length(nz1.rml)

```

[1] 22

```

> plot(check1.cv.rcal$est, xlim=c(0,p), ylim=c(-.3,.3),
+       xlab="", ylab="std diff", main="RCAL")
> abline(h=0)
> abline(h=max(abs(check1.cv.rcal$est)) *c(-1,1), lty=2)
> nz1.rcal <- which(late.cv.rcal$ips[[2]]$sel.bet[,1]!=0)-1
> text(nz1.rcal, -0.3, "x", cex=1)
> length(nz1.rcal)

```

[1] 6

```

> plot(fips1.cv.rml[iv==1], fips1.cv.rcal[iv==1], xlim=c(0,1), ylim=c(0,1),
+       xlab="RML", ylab="RCAL", main="fitted probabilities")
> abline(c(0,1))
> ## Computation of standardized calibration differences
> ##           within iv=0 group
> fips0.cv.rcal<-late.cv.rcal$mfp[,1]

```

```

> fips0.cv.rml <-late.cv.rml$mfpl[,1]
> check0.raw <- mn.ipw(x, 1-iv, fips.raw)
> check0.cv.rcal <- mn.ipw(x, 1-iv, fips0.cv.rcal)
> check0.cv.rml <- mn.ipw(x, 1-iv, fips0.cv.rml)
> par(mfrow=c(2,2))
> par(mar=c(4,4,2,2))
> plot(check0.raw$est, xlim=c(0,p), ylim=c(-.3,.3),
+       xlab="", ylab="std diff", main="Raw")
> abline(h=0)
> plot(check0.cv.rml$est, xlim=c(0,p), ylim=c(-.3,.3),
+       xlab="", ylab="std diff", main="RML")
> abline(h=0)
> abline(h=max(abs(check0.cv.rml$est)) *c(-1,1), lty=2)
> nz0.rml <- which(late.cv.rml$ips[[2]]$sel.bet[,1]!=0)-1
> text(nz0.rml, -0.3, "x", cex=1)
> length(nz0.rml)

```

[1] 22

```

> plot(check0.cv.rcal$est, xlim=c(0,p), ylim=c(-.3,.3),
+       xlab="", ylab="std diff", main="RCAL")
> abline(h=0)
> abline(h=max(abs(check0.cv.rcal$est)) *c(-1,1), lty=2)
> nz0.rcal <- which(late.cv.rcal$ips[[1]]$sel.bet[,1]!=0)-1
> text(nz0.rcal, -0.3, "x", cex=1)
> length(nz0.rcal)

```

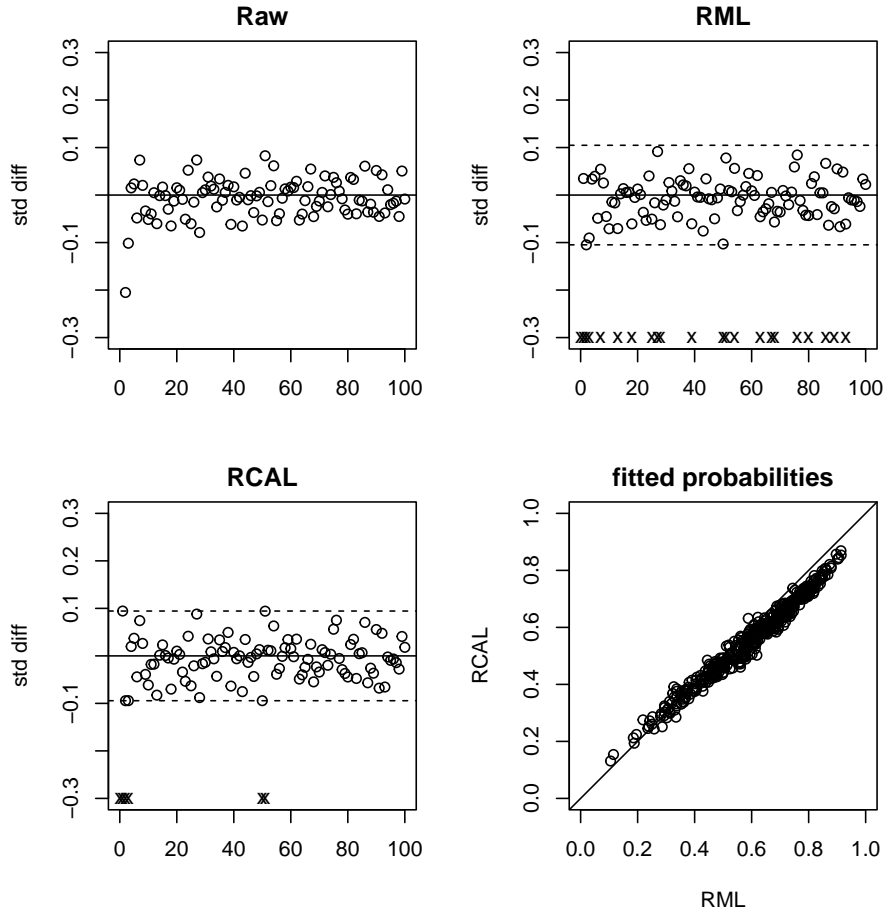
[1] 5

```

> plot(fips0.cv.rml[iv==0], fips0.cv.rcal[iv==0], xlim=c(0,1), ylim=c(0,1),
+       xlab="RML", ylab="RCAL", main="fitted probabilities")
> abline(c(0,1))

```

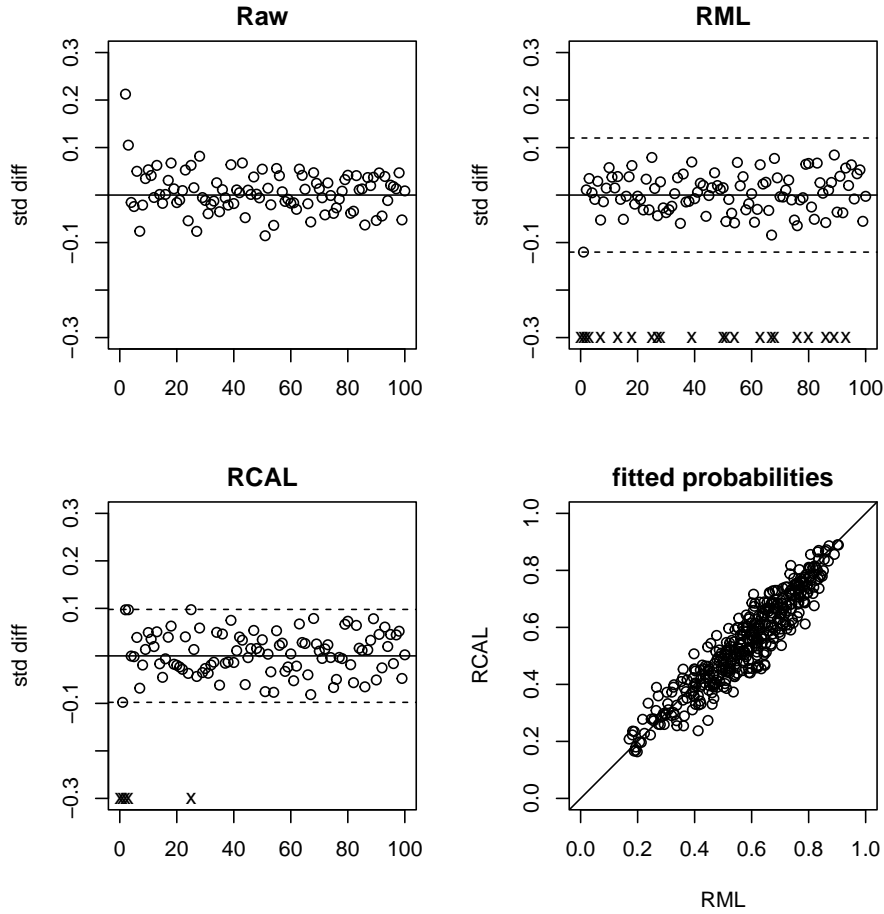
Figure 4: Standardized calibration differences and scatterplot of fitted IPS within the group $\{i : Z_i = 1, 1 \leq i \leq n\}$. Marks (x) are plotted at the indices corresponding to 6 or 22 nonzero coefficients out of 100 in total in the IPS model fitted using regularized calibrated estimation (RCAL) or regularized maximum likelihood estimation (RML) respectively.



From Figure 4 or 5, the maximum absolute standardized calibration difference from regularized calibrated estimation is noticeably smaller than that from regularized maximum likelihood estimation, which is moreover achieved with a smaller number of nonzero coefficients out of 100 in total in the fitted IPS model.

For further comparison, the following uses the function `late.regu.path` to compute fitted IPS over regularization paths. Figure 6 show how the maximum absolute standardized differences vary against the numbers of nonzero coefficients and relative

Figure 5: Standardized calibration differences and scatterplot of fitted IPS within the group $\{i : Z_i = 0, 1 \leq i \leq n\}$. Marks (x) are plotted at the indices corresponding to 5 or 22 nonzero coefficients out of 100 in total in the IPS model fitted using regularized calibrated estimation (RCAL) or regularized maximum likelihood estimation (RML) respectively.



variances in the $Z_i = 1$ group, similarly as in Tan (2020a), Section 6. Codes are also provided to produce corresponding plots in the $Z_i = 0$ group. In this example, it appears impossible for regularized maximum likelihood estimation to reduce the maximum absolute calibration difference to below 0.05, even as the Lasso penalty decreases and the number of nonzero coefficients and relative variance increase.

```
> set.seed(0)
> late.path.rcal <-
```

```

+ late.regu.path(fold=5*c(1,1,1), nrho=(1+10)*c(1,1,1), rho.seq=NULL,
+ y, tr, iv, fx=x, gx=x, hx=x, arm=1, d1=1, d2=3, ploss="cal", yloss="gaus")
> set.seed(0)
> late.path.rml <-
+ late.regu.path(fold=5*c(1,1,1), nrho=(1+10)*c(1,1,1), rho.seq=NULL,
+ y, tr, iv, fx=x, gx=x, hx=x, arm=1, d1=1, d2=3, ploss="ml", yloss="gaus")
> ## Computation of standardized calibration differences
> ## along regularization path within iv=1 group
> fips1.path.rcal <- late.path.rcal$mfp[[2]]
> mdiff1.path.rcal <- rep(NA, dim(fips1.path.rcal)[2])
> rvar1.path.rcal <- rep(NA, dim(fips1.path.rcal)[2])
> for (j in 1:dim(fips1.path.rcal)[2]) {
+ check1.path.rcal <- mn.ipw(x, iv, fips1.path.rcal[,j])
+ mdiff1.path.rcal[j] <- max(abs(check1.path.rcal$est))
+ rvar1.path.rcal[j] <-
+ var(1/fips1.path.rcal[iv==1,j])/mean(1/fips1.path.rcal[iv==1,j])^2
+ }
> fips1.path.rml <- late.path.rml$mfp[[2]]
> mdiff1.path.rml <- rep(NA, dim(fips1.path.rml)[2])
> rvar1.path.rml <- rep(NA, dim(fips1.path.rml)[2])
> for (j in 1:dim(fips1.path.rml)[2]) {
+ check1.path.rml <- mn.ipw(x, iv, fips1.path.rml[,j])
+ mdiff1.path.rml[j] <- max(abs(check1.path.rml$est))
+ rvar1.path.rml[j] <-
+ var(1/fips1.path.rml[iv==1,j])/mean(1/fips1.path.rml[iv==1,j])^2
+ }
> par(mfrow=c(1,2))
> par(mar=c(4,4,2,2))
> plot(late.path.rml$fips[[2]]$nz.all, mdiff1.path.rml,
+ xlim=c(0,p), ylim=c(0,.4), xlab="# nonzero", ylab="std diff")

```

```

> lines(late.path.rml$ips[[2]]$nz.all[!late.path.rml$ips[[2]]$non.conv],
+       mdiff1.path.rml, lty=3)
> points(late.path.rcal$ips[[2]]$nz.all[!late.path.rcal$ips[[2]]$non.conv],
+        mdiff1.path.rcal, pch=4)
> lines(late.path.rcal$ips[[2]]$nz.all[!late.path.rcal$ips[[2]]$non.conv],
+       mdiff1.path.rcal, lty=3)
> legend(120,.4, c("RML","RCAL"), pch=c(1,4), cex=.6)
> #
> plot(rvar1.path.rml, mdiff1.path.rml,
+      xlim=c(0,1), ylim=c(0,.4), xlab="rel var", ylab="std diff")
> lines(rvar1.path.rml, mdiff1.path.rml, lty=3)
> points(rvar1.path.rcal, mdiff1.path.rcal, pch=4)
> lines(rvar1.path.rcal, mdiff1.path.rcal, lty=3)
> legend(0.6,.4, c("RML","RCAL"), pch=c(1,4), cex=.6)

> set.seed(0)
> late.path.rcal <-
+ late.regu.path(fold=5*c(1,1,1), nrho=(1+10)*c(1,1,1), rho.seq=NULL,
+ y, tr, iv, fx=x, gx=x, hx=x, arm=0, d1=1, d2=3, ploss="cal", yloss="gaus")
> set.seed(0)
> late.path.rml <-
+ late.regu.path(fold=5*c(1,1,1), nrho=(1+10)*c(1,1,1), rho.seq=NULL,
+ y, tr, iv, fx=x, gx=x, hx=x, arm=0, d1=1, d2=3, ploss="ml", yloss="gaus")
> ## Computation of standardized calibration differences
> ##   along regularization path within iv=0 group
> fips0.path.rcal <- late.path.rcal$mfp[[1]]
> mdiff0.path.rcal <- rep(NA, dim(fips0.path.rcal)[2])
> rvar0.path.rcal <- rep(NA, dim(fips0.path.rcal)[2])
> for (j in 1:dim(fips0.path.rcal)[2]) {
+   check0.path.rcal <- mn.ipw(x, 1-iv, fips0.path.rcal[,j])
+   mdiff0.path.rcal[j] <- max(abs(check0.path.rcal$est))

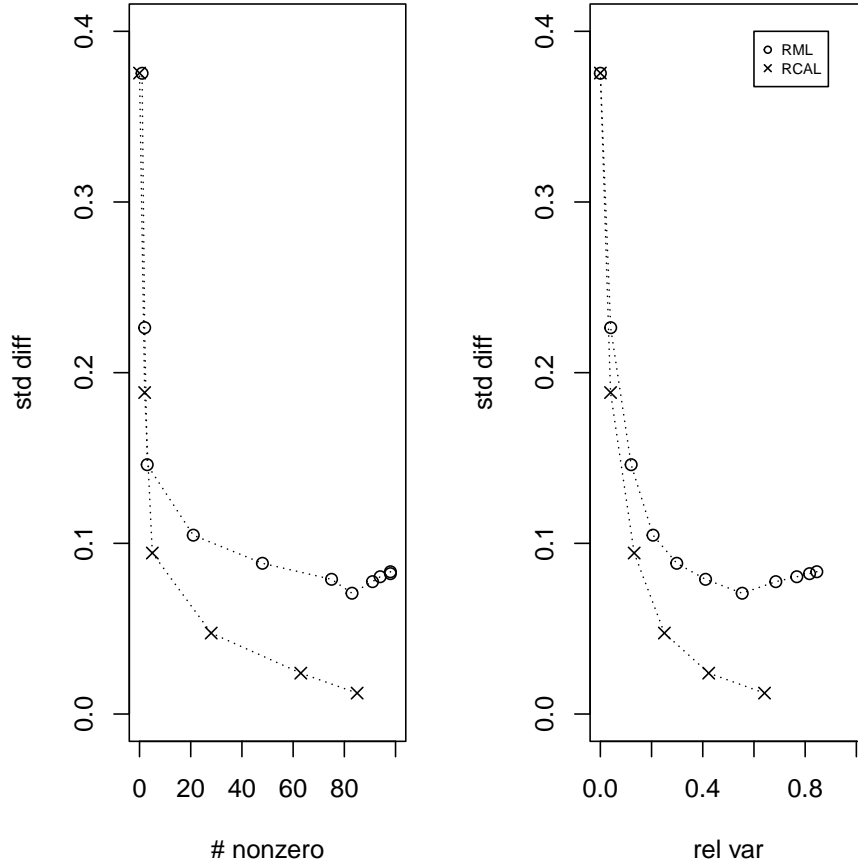
```

```

+   rvar0.path.rcal[j] <-
+   var(1/fips0.path.rcal[iv==0,j])/mean(1/fips0.path.rcal[iv==0,j])^2
+ }
> fips0.path.rml <- late.path.rml$mf[[1]]
> mdiff0.path.rml <- rep(NA, dim(fips0.path.rml)[2])
> rvar0.path.rml <- rep(NA, dim(fips0.path.rml)[2])
> for (j in 1:dim(fips0.path.rml)[2]) {
+   check0.path.rml <- mn.ipw(x, 1-iv, fips0.path.rml[,j])
+   mdiff0.path.rml[j] <- max(abs(check0.path.rml$est))
+   rvar0.path.rml[j] <-
+   var(1/fips0.path.rml[iv==0,j])/mean(1/fips0.path.rml[iv==0,j])^2
+ }
> par(mfrow=c(1,2))
> par(mar=c(4,4,2,2))
> plot(late.path.rml$ips[[2]]$nz.all, mdiff0.path.rml,
+       xlim=c(0,p), ylim=c(0,.4), xlab="# nonzero", ylab="std diff")
> lines(late.path.rml$ips[[2]]$nz.all[!late.path.rml$ips[[2]]$non.conv],
+       mdiff0.path.rml, lty=3)
> points(late.path.rcal$ips[[1]]$nz.all[!late.path.rcal$ips[[1]]$non.conv],
+        mdiff0.path.rcal, pch=4)
> lines(late.path.rcal$ips[[1]]$nz.all[!late.path.rcal$ips[[1]]$non.conv],
+        mdiff0.path.rcal, lty=3)
> legend(120,.4, c("RML","RCAL"), pch=c(1,4), cex=.6)
> plot(rvar0.path.rml, mdiff0.path.rml,
+       xlim=c(0,2.5), ylim=c(0,.4), xlab="rel var", ylab="std diff")
> lines(rvar0.path.rml, mdiff0.path.rml, lty=3)
> points(rvar0.path.rcal, mdiff0.path.rcal, pch=4)
> lines(rvar0.path.rcal, mdiff0.path.rcal, lty=3)
> legend(0.6,.4, c("RML","RCAL"), pch=c(1,4), cex=.6)

```


Figure 6: Maximum absolute standardized differences against the numbers of nonzero coefficients and relative variances within the group $\{i : Z_i = 1, 1 \leq i \leq n\}$.



REFERENCES

- Angrist, J.D., Imbens, G.W. and Rubin, D.B. (1996) Identification of causal effects using instrumental variables, *Journal of the American Statistical Association*, 91, 444–455.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. and Robins, J.M. (2018) Double/debiased machine learning for treatment and structural parameters, *The Econometrics Journal*, 21, C1–C68.
- Neyman, J. (1923) On the application of probability theory to agricultural experiments: Essay on principles, Section 9, translated in *Statistical Science*, 1990, 5,

465-480.

- Robins, J.M., Rotnitzky, A., and Zhao, L.P. (1994) Estimation of regression coefficients when some regressors are not always observed, *Journal of the American Statistical Association*, 89, 846-866.
- Rosenbaum, P.R. and Rubin, D.B. (1983) The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70, 41-55.
- Rubin, D.B. (1974) Estimating causal effects of treatments in randomized and non-randomized studies, *Journal of Educational Psychology*, 66, 688-701.
- Sun, B. and Tan, Z. (2020) High-dimensional model-assisted inference for local average treatment effects with instrumental variables, arXiv:2009.09286.
- Tan, Z. (2006) Regression and weighting methods for causal inference using instrumental variables, *Journal of the American Statistical Association*, 101, 1607-1618.
- Tan, Z. (2007) Comment: Understanding OR, PS and DR, *Statistical Science*, 22, 560-568.
- Tan, Z. (2020a) Regularized calibrated estimation of propensity scores with model misspecification and high-dimensional data, *Biometrika*, 107, 137-158.
- Tan, Z. (2020b) Model-assisted inference for treatment effects using regularized calibrated estimation with high-dimensional data, *Annals of Statistics*, 48, 811-837.
- Vytlačil, E. (2002) Independence, monotonicity, and latent index models: An equivalence result, *Econometrica*, 70, 331-341.