Chapter 6
Goodness-of-Fit Test for Generalized Linear Models

Goodness-of-fit (GOF) tests in regression analysis are mainly based on the observed residuals. In a series of articles, starting with Stute (1997), Stute established a general approach for GOF tests which is based on a marked empirical process (MEP), a standardized cumulative sum process obtained from the observed residuals. Resting upon the asymptotic limiting process of the MEP under the null hypothesis, Kolmogorov-Smirnov or Cramér-von Mises type tests can be stated as GOF tests. Their asymptotic distributions are derived through an application of the continuous mapping theorem. Since, in most cases, the asymptotic distributions depend on the model and, therefore, are not distribution free, further concepts are necessary to obtain the critical values for these tests.

In the literature, two approaches are discussed to handle the complicated structure of the limiting process of the MEP under the null hypothesis. In the first approach, the MEP is transformed in such a way that the resulting limiting process is a time-transformed Brownian motion with an assessable time transformation, compare Nikabadze and Stute (1997) and (Stute and Zhu, 2002). Originally, this technique was introduced by Khmaladze (1982) in the context of GOF tests based on estimated empirical processes. The second concept is based on the bootstrap, where the resampling scheme mimics the model under the null hypothesis. Resting upon the bootstrap data, the (bootstrap) MEP is derived. If one can show that this MEP tends to the same asymptotic process as the MEP of the original data does under the null hypothesis, the bootstrap MEP can be used to determine the critical value for the GOF statistic. Among others, this approach was used in Stute et al (1998) for parametric regression, in Dikta et al (2006) for binary regression, and in van Heel et al (2019) for multivariate binary regression models.

According to the general idea of bootstrap-based tests outlined in the introduction of Chap. 4, the bootstrap data has to be generated under the null hypothesis or close to it. If the asymptotic distribution of the bootstrapped statistic is the same as the corresponding one of the original data under the null hypothesis, critical values obtained from the bootstrap statistic can be used since they are derived under the null

G. Dikta and M. Scheer, *Bootstrap Methods*,
https://doi.org/10.1007/978-3-030-73480-0_6

hypothesis (the bootstrap data are generated under the null hypothesis) regardless whether the original data are following the null hypothesis or the alternative.

Denote, as usual, by $\mathbb{E}(Y \mid X = x)$ the regression function of Y at x (the conditional expectation of Y given $X = x$), and let

$$\mathscr{M} = \left\{ m(\beta^\top \cdot, \theta) \; : \; (\beta, \theta) \in \mathbb{R}^p \times \Theta \subset \mathbb{R}^{p+q} \right\}$$

define a parametric class based on a known function m. The general test problem for the GOF test is now

$$\mathbb{E}(Y \mid X = \cdot) \in \mathscr{M} \quad \text{versus} \quad \mathbb{E}(Y \mid X = \cdot) \notin \mathscr{M}.$$

Within the context of GLM with link function g, whether parametric or semi-parametric, there exists a $\beta_0 \in \mathbb{R}^p$ such that

$$\mathbb{E}(Y \mid X = x) = g^{-1}(\beta_0^\top x), \quad \text{for all } x \in \mathbb{R}^p.$$

Therefore, the MEP

$$\bar{R}_n^1 : \quad [-\infty, \infty] \ni u \longrightarrow \bar{R}_n^1(u) = n^{-1/2} \sum_{i=1}^{n} \big(Y_i - g^{-1}(\beta_n^\top X_i) \big) \mathrm{I}_{\{\beta_n^\top X_i \le u\}} \tag{6.1}$$

can be used for the original data and

$$R_n^{1*} : \quad [-\infty, \infty] \ni u \longrightarrow R_n^{1*}(u) = n^{-1/2} \sum_{i=1}^{n} \big(Y_i^* - g^{-1}(\beta_n^{*\top} X_i) \big) \mathrm{I}_{\{\beta_n^\top X_i \le u\}} \tag{6.2}$$

as bootstrap-based MEP, where the exact definition is given in Definitions 6.17 and 6.23, respectively.

In a parametric regression setup, where MEP-based statistics will be used for GOF tests, the following details will guarantee the validity of the bootstrap-based test:

1. Estimate the model parameter and build the MEP $\bar{R}_n^1$.
2. Determine the limit process of the MEP under the null hypothesis.
3. Generate bootstrap data according to the model, where the estimated parameters are used.
4. Repeat step (1) based on the bootstrap data and use R_n^{1*} as bootstrap-based MEP.
5. Verify that the bootstrap-based MEP tends to the limit process which is derived under (2).

The parameter estimation under (1) depends on the type of regression model. If we consider a semi-parametric GLM setup, LSE will be used to estimate the parameter since no further information about the distribution type of the error term is available. In this case, the wild bootstrap will be used. Otherwise, in the parametric GLM case, MLE will be applied and the bootstrap will be implemented parametrically.

Note that the indicator function in R_n^{1*} is based on β_n and not on β_n^*, as one would expect. This is mainly for performance reasons, see Remark 6.24. As shown in Sects. 6.4 and 6.5, both processes converge in distribution to the same centered Gaussian process $\bar{R}_\infty^1$ under the null hypothesis in the parametric and in the semi-parametric setup, if the appropriate assumptions can be guaranteed. Furthermore, the paths of $\bar{R}_\infty^1$ can be assumed to be continuous functions. Based on this result, Kolmogorov-Smirnov (D_n) and Cramér-von Mises (W_n^2) statistic are defined analogously to (4.7) and (4.8), respectively, by

$$D_n = \sup_{-\infty \leq t \leq \infty} \left| \bar{R}_n^1(t) \right|, \quad W_n^2 = n^{-1} \sum_{i=1}^{n} \left(\bar{R}_n^1(\beta_n^\top X_i) \right)^2.$$

Both statistics can be used to reveal discrepancies between the assumed model and the observations. By replacing $\bar{R}_n^1$ with R_n^{1*}, we get

$$D_n^* = \sup_{-\infty \leq t \leq \infty} \left| R_n^{1*}(t) \right|, \quad W_n^{*2} = n^{-1} \sum_{i=1}^{n} \left(R_n^{1*}(\beta_n^\top X_i) \right)^2,$$

the corresponding bootstrap statistics. Since both processes $\bar{R}_n^1$ and R_n^{1*} converge against the same Gaussian process under the null hypothesis, it follows, applying the continuous mapping theorem, that D_n and D_n^* as well as W_n^2 and W_n^{*2} also converge against the same limit distribution. But since the bootstrap data are always generated under the null hypothesis, we can now approximate the p-values of D_n and W_n^2 as usual by Monte Carlo application.

Our R-package `bootGOF` contains methods for performing the bootstrap tests we describe in this chapter. It is available on https://github.com/MarselScheer/bootGOF and CRAN. A brief introduction to the package can be found in the appendix. However, we deliberately do not use the `bootGOF`- package here because we want to illustrate how such complex resampling schemes can be implemented from scratch using simple (understandable) R-commands.

6.1 MEP in the Parametric Modeling Context

Usually modeling data is an iterative process where by fitting a model and investigating diagnostic aspects, like plots and test for assumptions, give ideas about potential improvements or serious misspecification. The GOF test based on the MEP is an additional tool that helps to detect if a fitted model contradicts the data one tries to model.

In this section, we apply the GOF test, based on the marked empirical process, to a real dataset in order to choose between a Poisson-, normal-, or negative-binomial model. Afterward, the GOF test is applied to artificial datasets in order to get a feeling for the test in situation where the truth is known.

Assume a parametric GLM with link function g. Under the notation stated in Definition 5.45, the following resampling scheme will be used for the GOF test.

Resampling Scheme 6.1

(A) *Calculate the MLE* $\hat{\beta}_n$ *and* $\hat{\phi}_n$ *for* $(Y_1, X_1), \ldots, (Y_n, X_n)$.
(B) *Obtain the MEP* (6.1) *and calculate* D_n *and/or* W_n^2 *accordingly.*
(C) *Set* $X^*_{\ell;i} = X_i$ *for all* $i = 1, \ldots, n$ *and all* $\ell = 1, \ldots, m$.
(D) *Generate* $Y^*_{\ell;i}$ *according to the density* $f(\cdot, \hat{\beta}_n, \hat{\phi}_n, X_i)$ *for all* $i = 1, \ldots, n$ *and all* $\ell = 1, \ldots, m$.
(E) *Calculate the MLE* $\hat{\beta}^*_{\ell;n}$ *and* $\hat{\phi}^*_{\ell;n}$ *based on* $(Y^*_{\ell;1}, X^*_{\ell;1}), \ldots, (Y^*_{\ell;n}, X^*_{\ell;n})$*, the MEP* $R^{*1}_{\ell;n}$ *according to* (6.2)*,* $D^*_{\ell;n}$ *and/or* $W^{*2}_{\ell;n}$*, for* $\ell = 1, \ldots, m$.
(F) *Determine the* $p-$*value of* D_n *within the simulated* $D^*_{\ell;n}$, $1 \leq \ell \leq m$ *and/or* W_n^2 *the* p*-value of* W_n^2 *within the simulated* $W^{*2}_{\ell;n}$, $1 \leq \ell \leq m$*, respectively.*

6.1.1 Implementation

First, we need the test statistic that will be resampled. The Cramér-von Mises test can be implemented as follows:

```
Rn1 <- function(mod, est_b_time_x) {

  # mod          - a model fit,
  # est_b_time_x - scalar product of the covariates and
  #                estimator of beta

  o_idx <- order(est_b_time_x)
  ordered_res <- residuals(mod, type = "response")[o_idx]
  dplyr::tibble(est_b_time_x = est_b_time_x[o_idx],
                res = ordered_res,
                Rn1_x = cumsum(ordered_res) / sqrt(length(o_idx)),
                ordering = o_idx)
}

CvM <- function(mod, est_b_time_x) {

  # mod          - a model fit,
  # est_b_time_x - scalar product of the covariates and
  #                estimator of beta

  Wn2 <- mean(Rn1(mod, est_b_time_x)$Rn1_x^2)
  Wn2
}
```

Note that this function itself uses the generic functions *predict* and *residuals* with a specific *type*-parameter. A fit created with “stats::glm” can be safely passed to the function because the corresponding functions “predict.glm” and “residuals.glm” respect the defined *type*-parameter. However, other packages can be used to fit a generalize linear model. Such packages usually provide their own set of *predict*- and *residuals*-function. In that case, the *type*-parameter of the corresponding package-specific function might have a different meaning or ignore the parameter completely, which then could lead to the wrong test statistic or an error message. Although “stats::glm” can fit various distributions, it does not offer the possibility to fit a negative-binomial model. One option, which works properly with our function for the test statistic, is “MASS::glm.nb”.

Next, we implement the Resampling Scheme 6.1. Fortunately, R provides a lot of infrastructure that allows an easy implementation.

```
gof_model_boot <- function(model, data, B = 1000) {

  # mod  - a model fit,
  #    + residuals(mod, type = "response") must return
  #      Y - m_est(X), where is the estimator of the
  #      regression function m
  #    + predict(model, type = "link") must return
  #      the scalar product of the covariates and
  #      estimator of beta
  #    + simulate(model) must generate generate
  #      target/dependent variables according to
  #      the fitted model
  # data - observed data
  # B    - number of bootstrapped MEPs

  # progress bar that appears if calculations will take more
  # than 1 second
  pb <- dplyr::progress_estimated(B, min_time = 1)

  est_b_time_x <- predict(model, type = "link")
  # Calculate the statistic for the original MEP
  Wn2 <- CvM(model, est_b_time_x = est_b_time_x)

  # copy to build up the bootstrap data set
  data_boot <- data

  # name of the target/dependent variable
  y_name <- all.vars(formula(model), max.names = 1)

  # bootstrap the statistic
  Wn2_boot <- sapply(seq_len(B), function(i) {

    pb$tick()$print() # print progress

    # due to Step C only the target/dependent variable
    # needs to be updated.
    data_boot[[y_name]] <- simulate(model)[,1]
```

```
    # refit the model using the bootstrapped data set
    m_boot <- update(model, formula. = formula(model),
                     data = data_boot)

    # Calculate the statistic for the bootstrapped MEP
    CvM(m_boot, est_b_time_x)
  })
  ret <- list(Wn2_boot = Wn2_boot,
              Wn2 = Wn2,
              pvalue_cvm = mean(Wn2_boot > Wn2))
  ret
}
```

The following is only a convenient function for displaying the estimated marked empirical process and model residuals while also showing bootstrapped versions.

```
plot_Rn1_and_residuals <- function(model, data, B) {
  # mod  - a model fit,
  #   + residuals(mod, type = "response") must return
  #     Y - m_est(X), where is the estimator of the
  #     regression function m
  #   + predict(model, type = "link") must return
  #     the scalar product of the covariates and
  #     estimator of beta
  #   + simulate(model) must generate generate
  #     target/dependent variables according to
  #     the fitted model
  # data - observed data
  # B    - number of bootstrapped MEPs

  y_name <- all.vars(formula(model), max.names = 1)
  est_b_time_x <- predict(model, type = "link")
  # MEP of the original model
  org_model <- Rn1(model, est_b_time_x) %>% dplyr::as_tibble()

  # bootstrapped MEP
  boot_model <- purrr::map_dfr(seq_len(B), function(boot_idx){
    # due to Step C only the target/dependent variable
    # needs to be updated.
    data[[y_name]] <- simulate(model, data = data)[,1]

    # refit the model using the bootstrapped data set
    # and calculate the MEP
    update(model, formula. = formula(model), data = data) %>%
      Rn1(est_b_time_x = est_b_time_x) %>%
      dplyr::as_tibble() %>%
      dplyr::mutate(original = FALSE, idx = boot_idx)
  })

  # statistics for the bootstrapped models
  Wn2_boot <- boot_model %>%
    dplyr::group_by(idx) %>%
```

```
    dplyr::summarise(CvM = mean(Rn1_x^2))

  pvalue_cvm <- mean(Wn2_boot$CvM > mean(org_model$Rn1_x^2))

  plot_Rn1 <- boot_model %>%
    ggplot(aes(x = est_b_time_x, y = Rn1_x)) +
    geom_line(aes(group = idx), alpha = 0.1) +
    geom_line(data = org_model, color = "red") +
    ggtitle(paste0("p-value (CvM) = ", pvalue_cvm))

  plot_res <- boot_model %>%
    ggplot(aes(x = est_b_time_x, y = res)) +
    geom_point(alpha = 0.1) +
    geom_point(data = org_model, color = "red")

  cowplot::plot_grid(plot_Rn1, plot_res, nrow = 2)
}
```

Similar as before, this function uses generic functions, namely, "simulate" and "update". A fit created with "stats::glm" or "MASS::glm.nb" can be safely passed to this function. If another package is used, one should check that "simulate" really simulates the dependent variable and "update" refits the model using the generated dataset.

6.1.2 Bike Sharing Data

In Sect. 5.3, we prepared and analyzed the ridership data, which resulted in four model candidates. The corresponding diagnostic plots were already presented and briefly discussed in that section. Here, we apply the bootstrap-based goodness-of-fit test to obtain another indicator for inappropriate models.

As a reminder, we briefly repeat the steps from Sect. 5.3, i.e., import and preprocess/wrangle the dataset and subset it to the dates before hurricane "Sandy":

```
ridership <- readr::read_csv("day.csv") %>%
  data_preprocess()

## Parsed with column specification:
## cols(
##   instant = col_double(),
##   dteday = col_date(format = ""),
##   season = col_double(),
##   yr = col_double(),
##   mnth = col_double(),
##   holiday = col_double(),
##   weekday = col_double(),
##   workingday = col_double(),
##   weathersit = col_double(),
##   temp = col_double(),
```

```
##   atemp = col_double(),
##   hum = col_double(),
##   windspeed = col_double(),
##   casual = col_double(),
##   registered = col_double(),
##   cnt = col_double()
## )

ridership <-
  ridership %>%
  dplyr::filter(dteday < lubridate::ymd("2012-10-29"))
```

In order to get an idea how the upcoming plots of the residuals and the estimated marked empirical process would look like if the model is correct, we take the ridership data and generate the target according to a fitted model and then apply the GOF test, see Fig. 6.1.

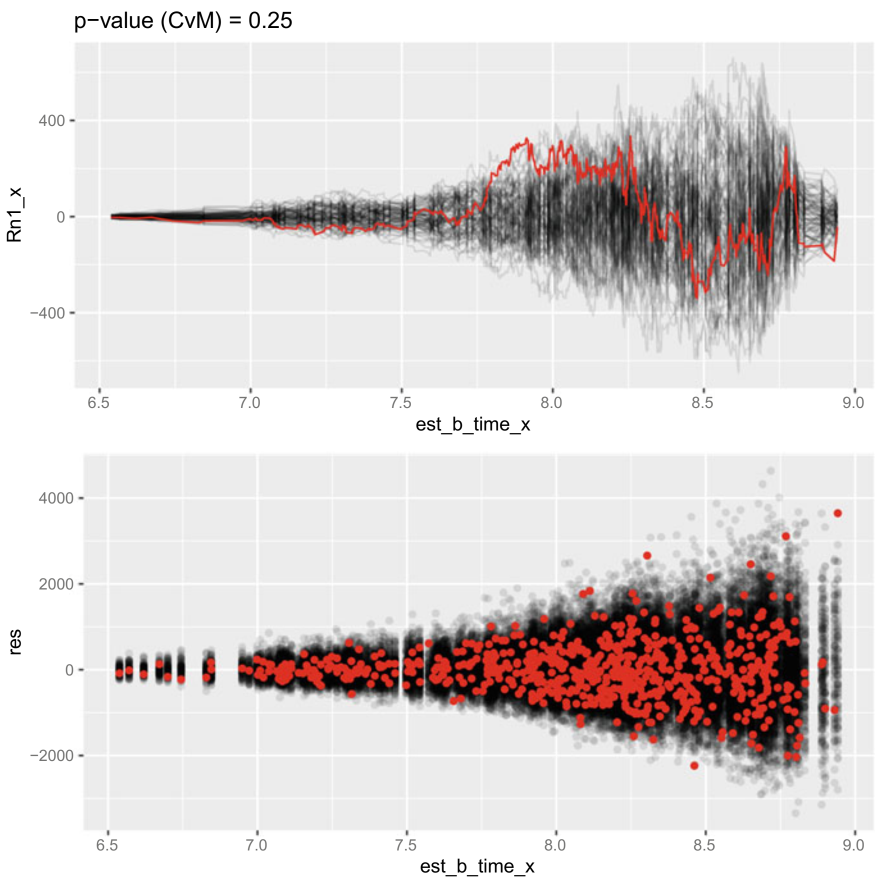

Fig. 6.1 Estimated marked empirical process and residuals in red and the corresponding bootstrapped versions in black if the ridership data would follow a negative-binomial model

```
frml <- y ~ temp + I(temp^2) + hum_imp + I(hum_imp^2) +
   windspeed + yr*season + workingday +
   weathersit + holiday + christmas
fit_nb <- MASS::glm.nb(frml, data = ridership)
ridership_generated <- ridership
# generate riderships that follow a negative-binomial
# distribution according to the fitted model
set.seed(123,kind ="Mersenne-Twister",normal.kind ="Inversion")
ridership_generated$y <- simulate(fit_nb, data = data)[,1]
fit_nb_generated <- MASS::glm.nb(frml, data = ridership_generated)
plot_Rn1_and_residuals(fit_nb_generated,
                       data = ridership_generated, B = 100)
```

Obviously, the estimated marked empirical process in Fig. 6.1 does not show more extreme behavior than the 100 bootstrapped versions and the residuals show a similar pattern as the residuals of the 100 bootstrapped model fits.

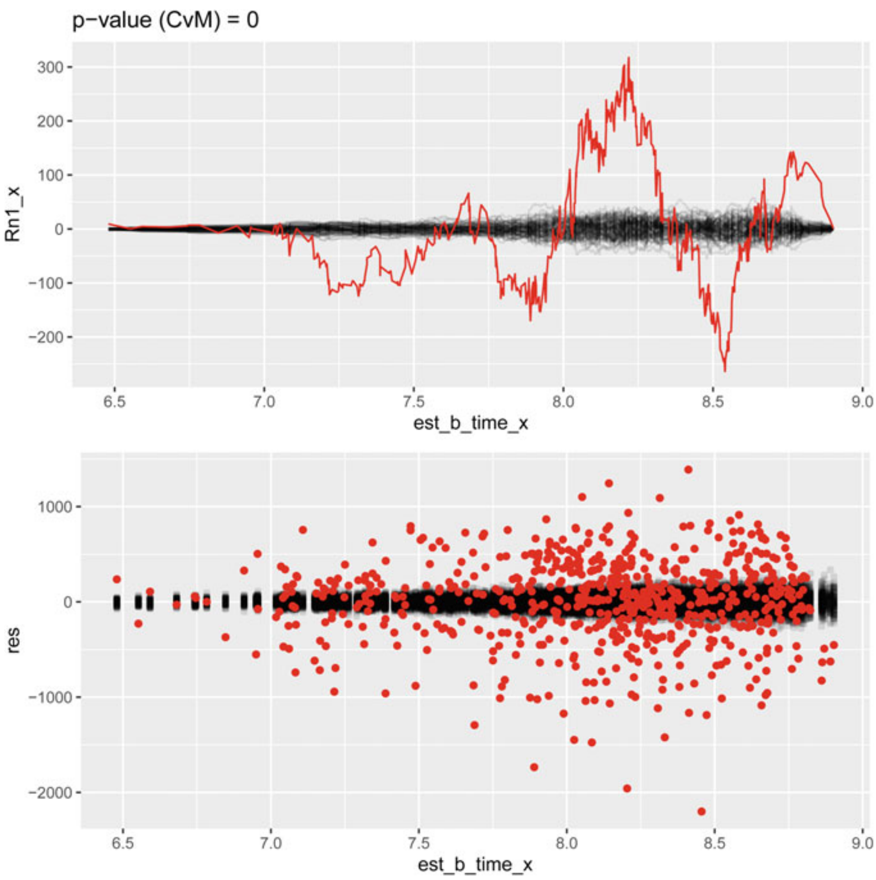

Fig. 6.2 Estimated marked empirical process and residuals in red and the corresponding bootstrapped versions in black for the Poisson model

Now, we apply the goodness-of-test to the four model candidates from the previous section. We start with the Poisson model, see Sect. 5.3 for the model output.

```
fit_poi <- glm(frml, data = ridership, family = poisson())
set.seed(123,kind ="Mersenne-Twister",normal.kind ="Inversion")
plot_Rn1_and_residuals(fit_poi, data = ridership, B = 100)
```

Figure 6.2 reveals that the estimated marked empirical process (as well as the residuals) in the observed data behave very differently than its bootstrapped version. This results in rejecting the corresponding null hypothesis of the GOF test for our fitted Poisson model. Fitting a quasi-Poisson model is possible but the parametric bootstrap is not possible because the distribution is not fully defined. Therefore, for the quasi-Poisson model one would have to use other diagnostic checks.

```
fit_qpoi <- glm(frml, data = ridership, family = quasipoisson())
```

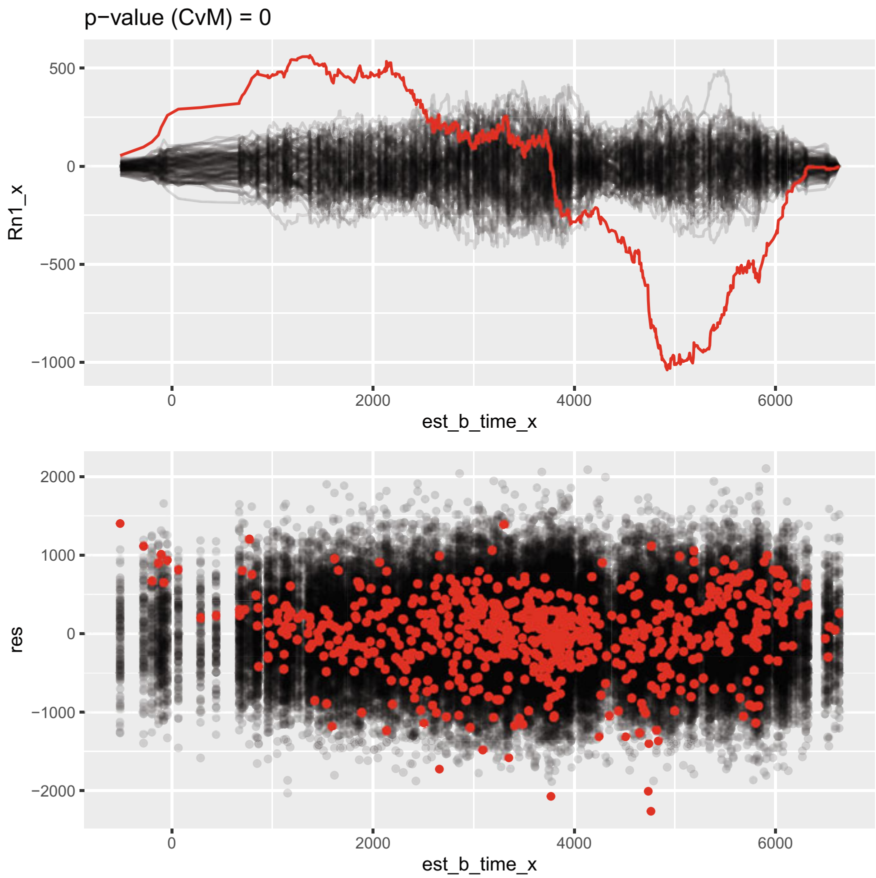

Fig. 6.3 Estimated marked empirical process and residuals in red and the corresponding bootstrapped versions in black for the Gaussian model

Trying the normal distributions without log-transformation also reveals that the estimated marked empirical process behaves very differently as its bootstrapped version, see Fig. 6.3.

```
fit_norm <- glm(frml, data = ridership, family = gaussian())
set.seed(123,kind ="Mersenne-Twister",normal.kind ="Inversion")
plot_Rn1_and_residuals(fit_norm, data = ridership, B = 100)
```

As in the last section, the normal distribution with log-transformations behaves surprisingly well, though the residuals seem to behave differently compared to the bootstrapped residuals, see Fig. 6.4.

```
fit_lognorm <-
 ridership %>%
 mutate(y = log(y)) %>%
 glm(frml, data = ., family = gaussian())
set.seed(123,kind ="Mersenne-Twister",normal.kind ="Inversion")
plot_Rn1_and_residuals(fit_lognorm, data = ridership, B = 100)
```

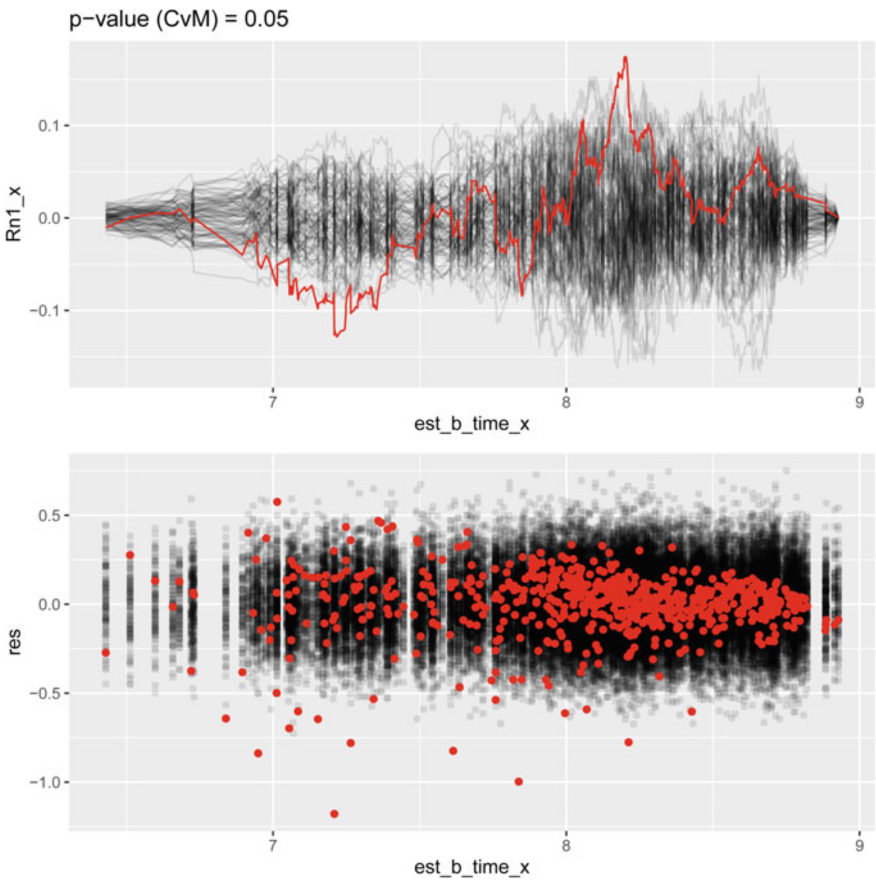

Fig. 6.4 Estimated marked empirical process and residuals in red and the corresponding bootstrapped versions in black for the Gaussian model with log-transformed target

Finally, using the negative-binomial distribution shows not too much deviations of the estimated marked empirical process from its bootstrapped versions, see Fig. 6.5. However, the residuals at the right end of the plot seem to indicate that the bootstrapped residuals have larger variance compared to the variance of the residuals based on the original data.

```
fit_nb <- MASS::glm.nb(frml, data = ridership)
set.seed(123,kind ="Mersenne-Twister",normal.kind ="Inversion")
plot_Rn1_and_residuals(fit_nb, data = ridership, B = 100)
```

Anyway, at this point, one would probably decide to go on with the negative-binomial model maybe also with the log-transformed Gaussian model (if the diagnostic checks for the quasi-Poisson model also indicate that it does not fit the data well) and start investigating the low residuals that seem to stand apart from the bootstrapped residuals as well as try to find the root cause for the smaller variance.

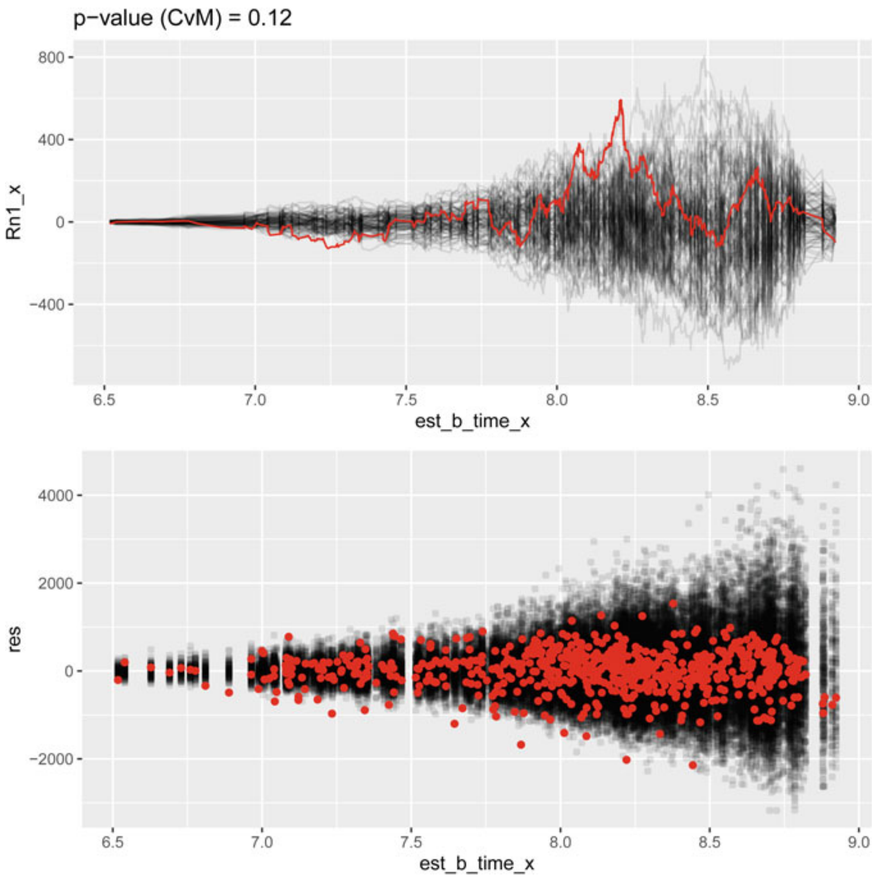

Fig. 6.5 Estimated marked empirical process and residuals in red and the corresponding bootstrapped versions in black for the negative-binomial model

6.1.3 Artificial Data

By generating the datasets we are able to judge if the result of the GOF test is correct. This provides at least some ideas about the limits of the GOF test. Of course, the simulations in this section are far from being exhaustive and, for instance, changing a β of the true model or the distribution of the covariates may lead to different results. Furthermore, a practitioner probably wants to investigate the GOF test himself in a specific situation when he has created reasonable model candidates. Consider the bike sharing example from the last section. It probably makes more sense in that particular case to take the negative-binomial model that passed the GOF test and artificially introduce additional covariates, for instance, a squared term, simulate the outcome variable and then check whether and when the GOF test is able to detect that a model without that new covariate is misspecified or just to get an idea of what we could expect if the model would be correct like Fig. 6.1.

We use a simple linear (Gaussian) model

$$Y = \beta_1 X_1^2 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon,$$

where X_1 and X_2 are uniformly distributed, X_3 is Bernoulli distributed, see Fig. 6.6.

```
genData <- function(N, coef_x1_square, coef_x2, coef_x3) {
  # N - sample size
  # data is generated according to
  # normal distribution with variance one
  # and mean
  # 10 + coef_x1_square * X1^2 + coef_x2 * X2 + coef_x3 * X3

  d <- data.frame(
    X1 = runif(N, 0, 3),
    X2 = runif(N, 1, 2),
    X3 = rbinom(N, size = 1, prob = 0.3),
    noise1 = runif(N),
    noise2 = runif(N)
  )
  lin_comb <- 10 + coef_x1_square * d$X1^2 +
    coef_x2 * d$X2 +
    coef_x3 * d$X3
  d$X3 <- as.factor(d$X3)
  d$Y <- rnorm(N, mean = lin_comb, sd = 1)
  return(d)
}
set.seed(123,kind ="Mersenne-Twister",normal.kind ="Inversion")
gaus_data <- genData(200, coef_x1_square = 1, coef_x2 = 2,
                     coef_x3 = 3)
```

```
GGally::ggpairs(gaus_data[, c("X1", "X2", "X3", "Y")])
```

One way to approach this dataset is to start with a backward selection.

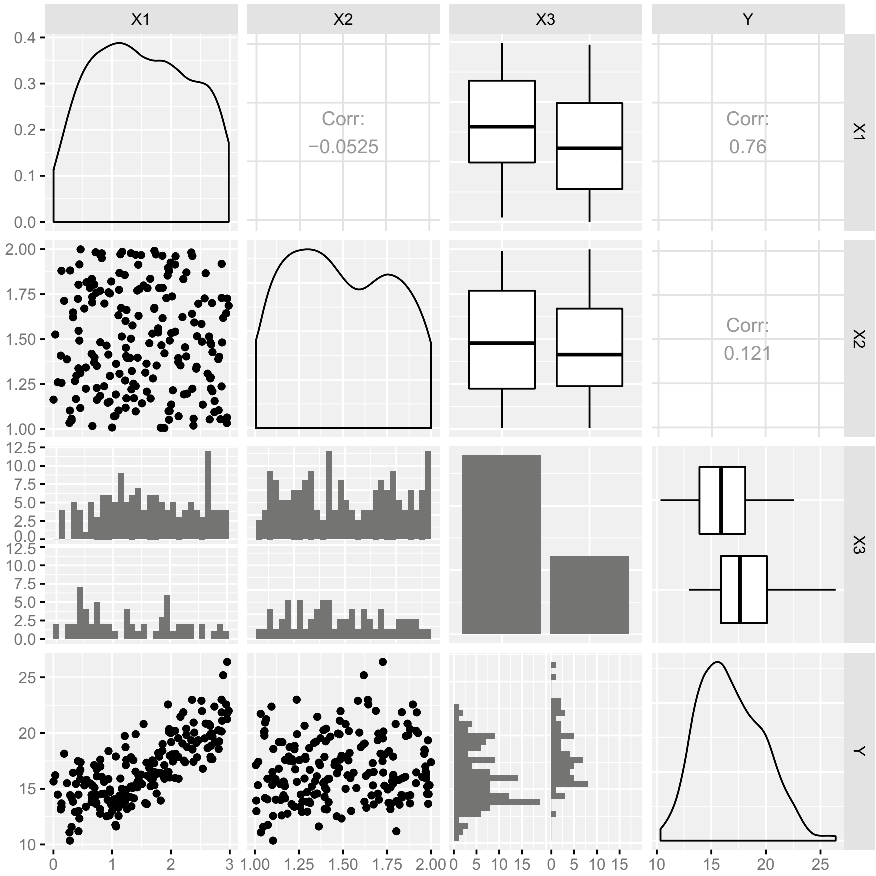

Fig. 6.6 Artificial dataset following the model $Y = \beta_1 X_1^2 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

```
fit <- glm(formula = Y ~ X1 + X2 + X3 + noise1 + noise2,
    data = gaus_data, family = "gaussian")
(fit)
```

```
##
## Call: glm(formula = Y ~ X1 + X2 + X3 + noise1 + noise2,
## family = "gaussian",
##     data = gaus_data)
##
## Coefficients:
## (Intercept)          X1          X2         X31
##      8.3225      3.0601      1.9339      3.2109
##      noise1      noise2
```

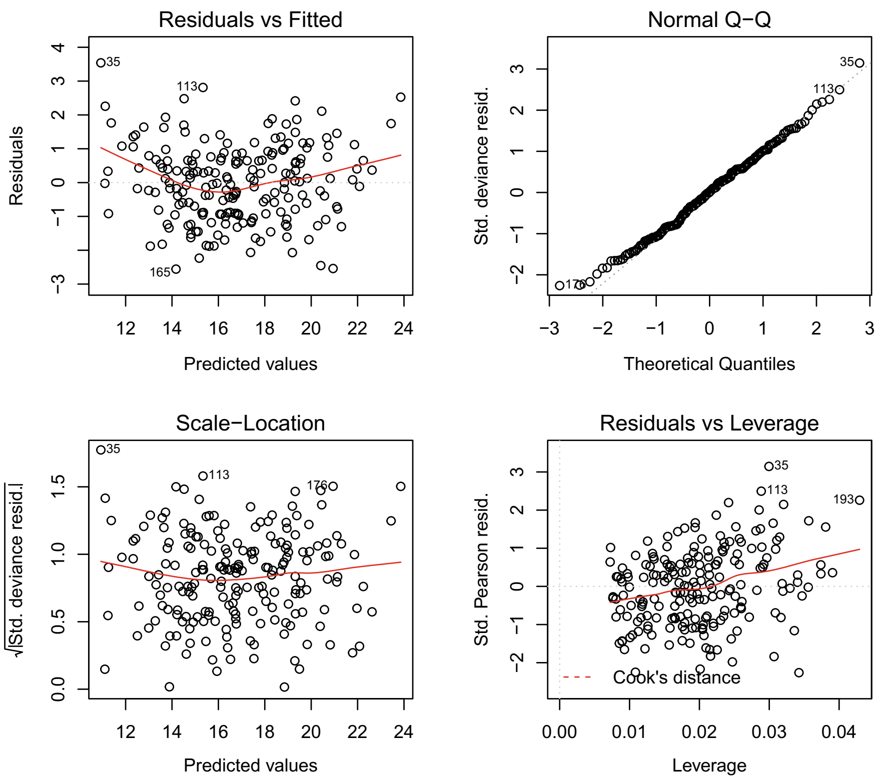

Fig. 6.7 Diagnostic plots for a linear model that only incorporate first-degree terms for X_1, X_2, and X_3

```
##       0.1959      -0.3021
##
## Degrees of Freedom: 199 Total (i.e. Null);  194 Residual
## Null Deviance:        1699
## Residual Deviance: 253.8     AIC: 629.3
```

This rules out the noise variables and the usual diagnostic plots already look quite promising, see Fig. 6.7.

```
fit <- glm(formula = Y ~ X1 + X2 + X3,
           data = gaus_data, family = "gaussian")
par(mfrow = c(2,2))
plot(fit)
```

However, the GOF test rejects the model, see Fig. 6.8. Though that figure does not indicate how to improve the model the residual plot in Fig. 6.7 indicates non-linearity.

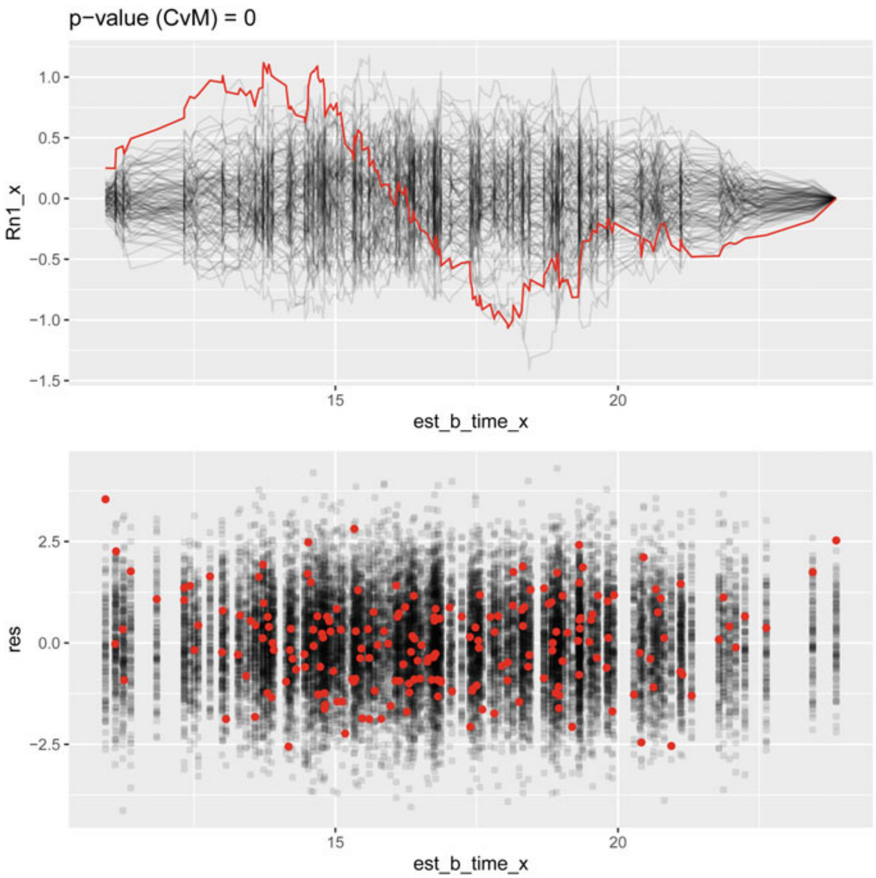

Fig. 6.8 Estimated marked empirical process and residuals in red and the corresponding bootstrapped versions in black for a linear model that only incorporates first-degree terms for X_1, X_2, and X_3

```
plot_Rn1_and_residuals(fit, gaus_data, B = 100)
```

Therefore, we try second-order terms

```
fit <- glm(formula = Y ~ X1 + X2 + X3 + I(X1^2) + I(X2^2) +
             X1:X2 + X1:X3 + X2:X3,
           data = gaus_data, family = "gaussian")
step(fit)


  ## Start:  AIC=576.44
  ## Y ~ X1 + X2 + X3 + I(X1^2) + I(X2^2) + X1:X2 + X1:X3 + X2:X3
  ##
```

```
##            Df Deviance    AIC
## - X1:X2     1   189.18 574.45
## - X2:X3     1   189.28 574.56
## - I(X2^2)  1   189.36 574.65
## - X1:X3     1   190.19 575.51
## <none>          189.17 576.44
## - I(X1^2)  1   254.21 633.54
##
## Step:  AIC=574.45
## Y ~ X1 + X2 + X3 + I(X1^2) + I(X2^2) + X1:X3 + X2:X3
##
##            Df Deviance    AIC
## - X2:X3     1   189.28 572.56
## - I(X2^2)  1   189.38 572.66
## - X1:X3     1   190.19 573.52
## <none>          189.18 574.45
## - I(X1^2)  1   254.30 631.61
##
## Step:  AIC=572.56
## Y ~ X1 + X2 + X3 + I(X1^2) + I(X2^2) + X1:X3
##
##            Df Deviance    AIC
## - I(X2^2)  1   189.50 570.79
## - X1:X3     1   190.31 571.64
## - X2        1   190.79 572.15
## <none>          189.28 572.56
## - I(X1^2)  1   254.72 629.95
##
## Step:  AIC=570.79
## Y ~ X1 + X2 + X3 + I(X1^2) + X1:X3
##
##            Df Deviance    AIC
## - X1:X3     1   190.50 569.84
## <none>          189.50 570.79
## - I(X1^2)  1   255.20 628.32
## - X2        1   268.75 638.67
##
## Step:  AIC=569.84
## Y ~ X1 + X2 + X3 + I(X1^2)
##
##            Df Deviance    AIC
## - X1        1   190.82 568.17
## <none>          190.50 569.84
## - I(X1^2)  1   255.89 626.86
```

```
## - X2         1    271.00 638.34
## - X3         1    582.17 791.27
##
## Step:  AIC=568.17
## Y ~ X2 + X3 + I(X1^2)
##
##            Df Deviance    AIC
## <none>          190.82 568.17
## - X2         1   273.42 638.11
## - X3         1   584.96 790.22
## - I(X1^2)  1  1471.54 974.72
##
## Call: glm(formula = Y ~ X2 + X3 + I(X1^2), family =
## "gaussian", data = gaus_data)
##
## Coefficients:
## (Intercept)           X2           X31       I(X1^2)
##      9.6133       2.2093       3.0953        0.9832
##
## Degrees of Freedom: 199 Total (i.e. Null);  196 Residual
## Null Deviance:       1699
## Residual Deviance: 190.8     AIC: 568.2
```

The diagnostic plots for the resulting model show that the non-linearity was reduced, see Fig. 6.9 and also the GOF test does not reject the new model, see Fig. 6.10.

```
fit <- glm(formula = Y ~ I(X1^2) + X2 + X3,
           data = gaus_data, family = "gaussian")
par(mfrow = c(2,2))
plot(fit)
```

```
plot_Rn1_and_residuals(fit, gaus_data, B = 100)
```

In this particular situation, the GOF test clearly rejected our first model, while the diagnostic plots only slightly indicated that the model is not correct. One should be aware of the fact that this might be also the other way around. In order to illustrate this, we generate a dataset, where the Bernoulli-distributed variable has a larger impact. The diagnostic plots make it obvious that the model is misspecified, see Fig. 6.11, but the GOF test is not able to detect that because this drastically increases the variance of the bootstrapped MEP, see Fig. 6.12.

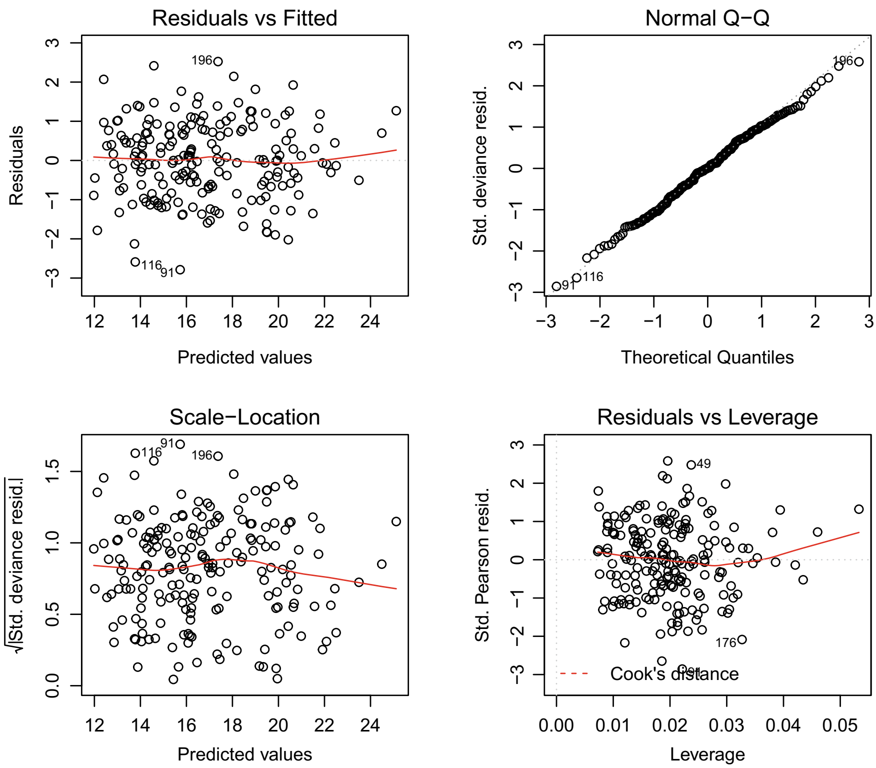

Fig. 6.9 Diagnostic plots for a linear model that incorporate all terms of the true model

```
set.seed(123,kind ="Mersenne-Twister",normal.kind ="Inversion")
gaus_data2 <- genData(200, coef_x1_square = 1, coef_x2 = 2,
                      coef_x3 = 6)
fit <- glm(formula = Y ~ I(X1^2) + X2, data = gaus_data2,
           family = "gaussian")
```

```
plot_Rn1_and_residuals(fit, gaus_data2, B = 100)
```

In this particular situation and this particular example, the GOF test rejected the model without X_1^2. In order to get a feeling for how reproducible this outcome would be, Fig. 6.13 shows the results of a small simulation study. Furthermore, in that simulation study, we add a model that misses term X_3 to see the performance of the GOF test with respect to this alternative.

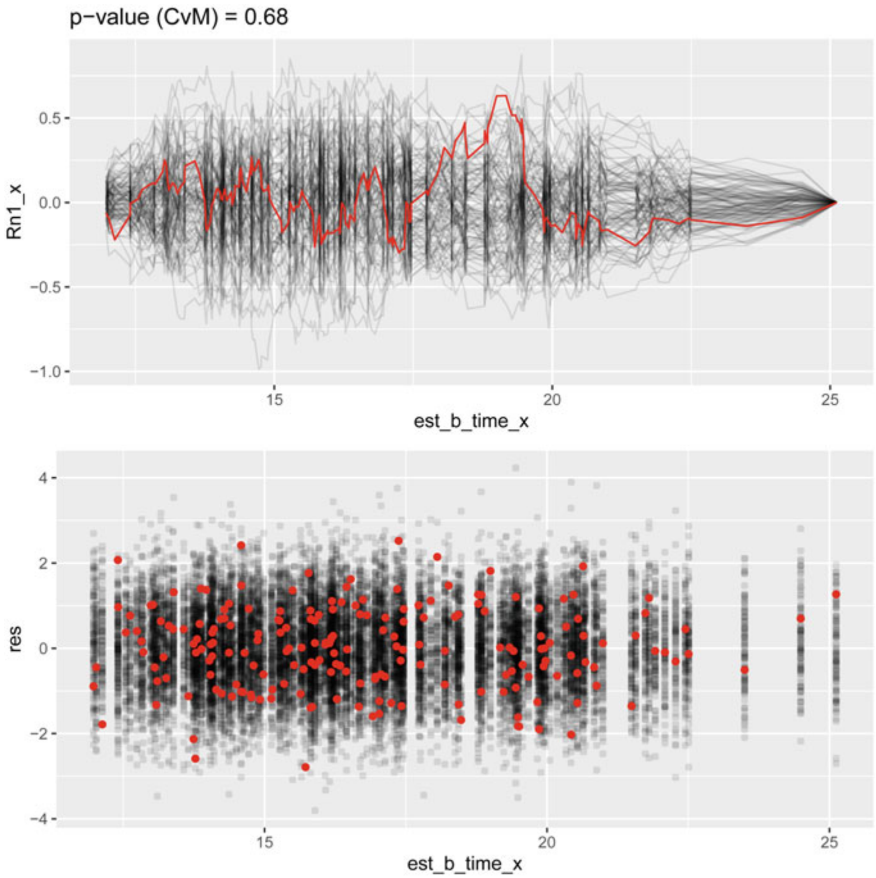

Fig. 6.10 Estimated marked empirical process and residuals in red and the corresponding bootstrapped versions in black for a linear model that incorporate all terms of the true model

```
gof_boot <- function(data, formula_str) {
 # fits a guassian model, performs
 # parametric GOF test and returns
 # the corresponding p-values

 # data - original data set
 # formula_str - a formula as a string

 frml <- as.formula(formula_str)
 m <- glm(frml, data = data, family = gaussian())

 gof <- gof_model_boot(m, data, B = 100)
 gof$pvalue_cvm
}
```

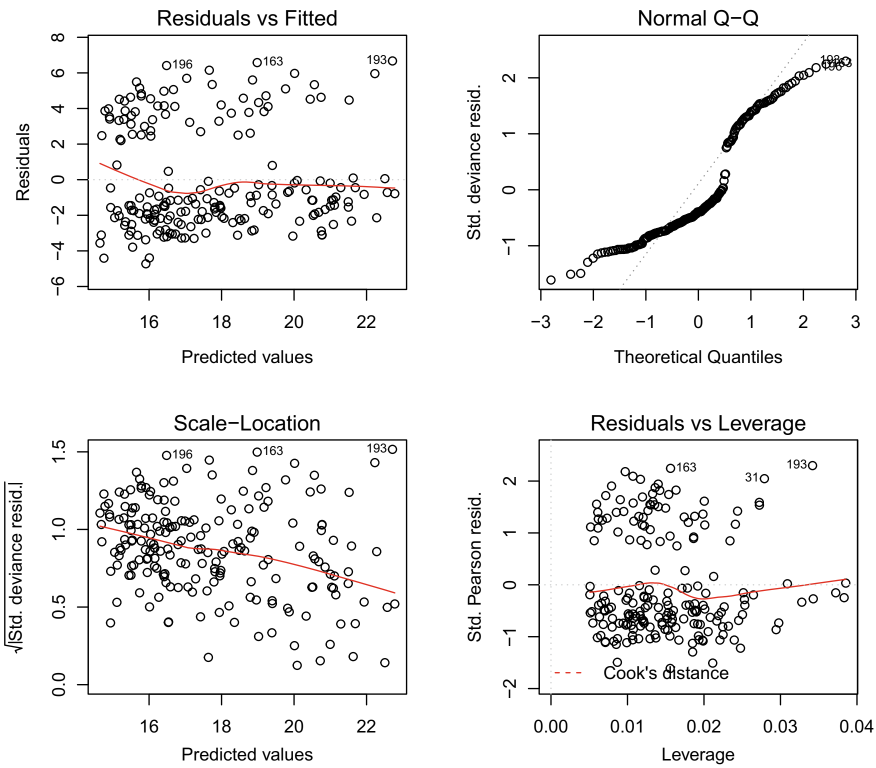

Fig. 6.11 Diagnostic plots for a linear model that incorporates all terms of the true model besides X_3

```
dg <- simTool::expand_tibble(
 proc = "genData",
 N = 200,
 coef_x1_square = 1,
 coef_x2 = 2,
 coef_x3 = 6)
pg <- simTool::expand_tibble(
 fun = c("gof_boot"),
 formula_str = c("Y ~ X1 + X2 + X3",
                 "Y ~ I(X1^2) + X2")
 )

eg <- simTool::eval_tibbles(
 data_grid = dg, proc_grid = pg,
 replications = 100, ncpus = 3,
 cluster_global_objects = ls())
```

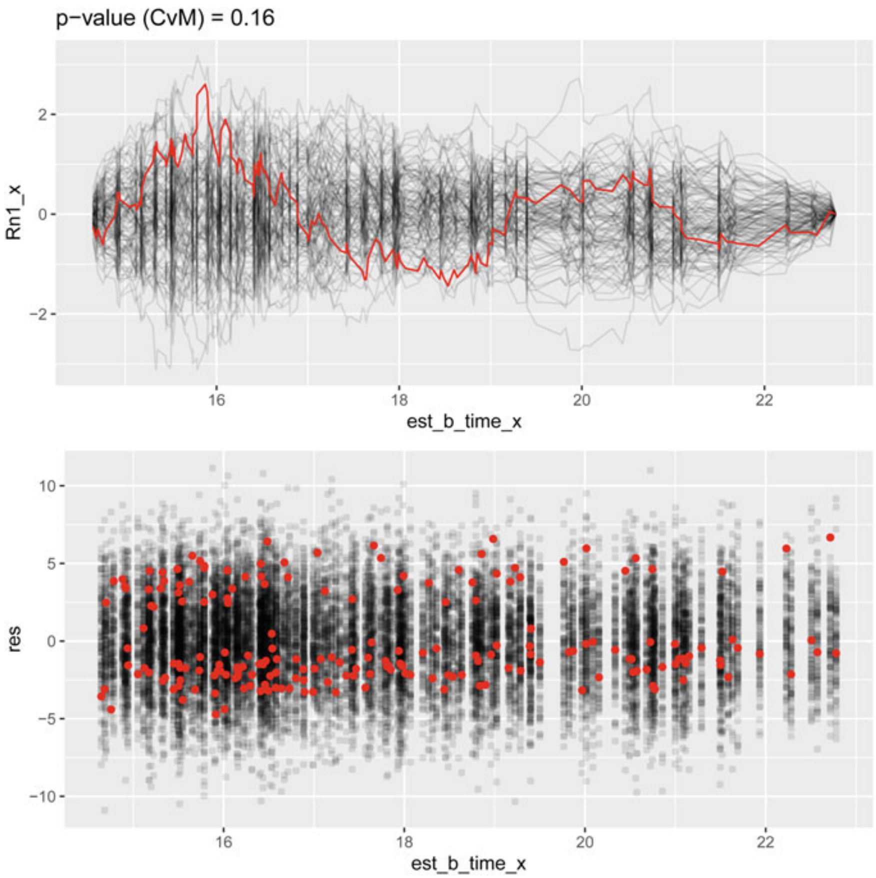

Fig. 6.12 Estimated marked empirical process and residuals in red and the corresponding bootstrapped versions in black for a linear model that incorporates all terms of the true model besides X_3

```
eg$simulation %>%
 ggplot(aes(x = results, color = formula_str)) +
 stat_ecdf() +
 geom_abline(slope = 1, intercept = 0) +
 #facet_grid(formula_str ~ N) +
 theme(legend.position = "top")
```

As one can see from Fig. 6.13, the GOF test rejects the model with missing X_1^2 with high probability but has a hard time if X_3 is not part of the model. But excluding X_3 from the model if it has such a large impact and then applying the GOF test makes no sense. From that point of view, this aspect of the simulation makes no sense. However, X_3 might not have been recorded during the creation of the dataset. In such a case, it would not be possible to include X_3 in the model and the GOF test

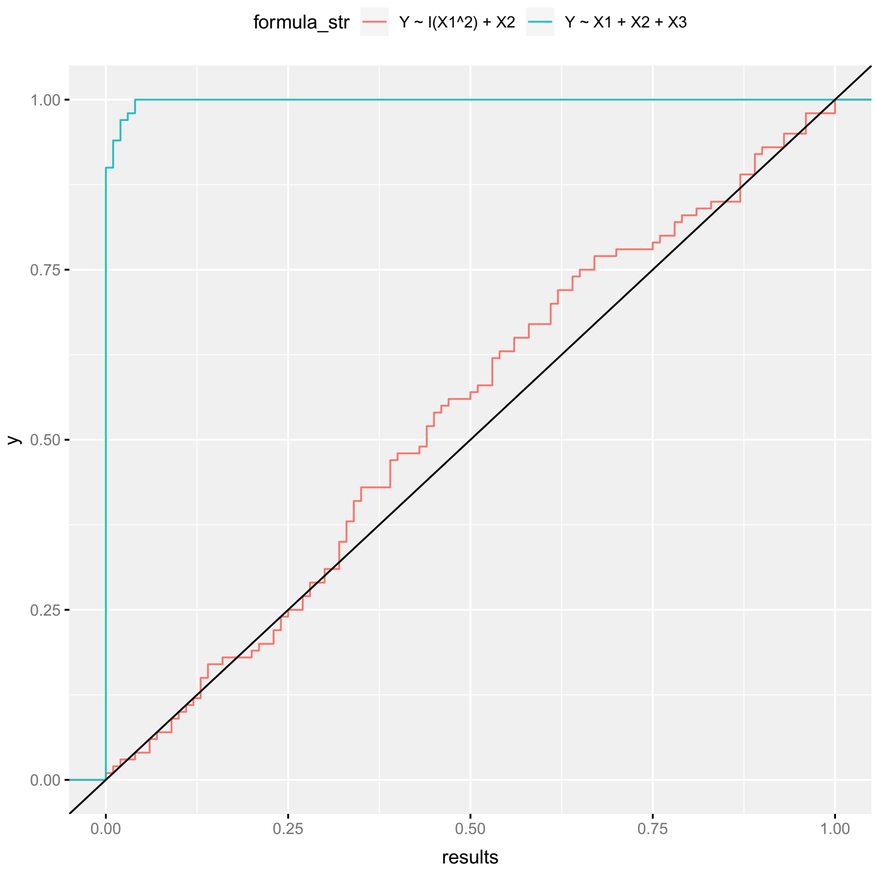

Fig. 6.13 Empirical cumulative distribution function of the p-values of the parametric GOF test

has only limited power to detect this. This shows that it makes it necessary to still consult other tests and plots to get an overall picture of misspecifications.

6.2 MEP in the Semi-parametric Modeling Context

In Sect. 5.4, it was assumed that

$$Y = m(X, \vartheta) + \varepsilon.$$

In the GLM context, this is specialized to

$$Y = m(\beta^\top X) + \varepsilon,$$

compare Definition 5.63.

The procedure for performing the GOF test is the same as the GOF test under the parametric GLM. Only the parameter estimations are no longer done with MLE but with LSE. Also the resampling is no longer performed according to a given distribution model but with the wild bootstrap. This results in the following resampling scheme.

Resampling Scheme 6.2

(A) *Calculate the LSE* $\hat{\beta}_n$ *based on* $(Y_1, X_1), \ldots, (Y_n, X_n)$.
(B) *Determine the estimated residuals* $\hat{\varepsilon}_i = Y_i - m(\beta_n^\top X_i)$, *for* $i = 1, \ldots, n$.
(C) *Obtain the MEP* (6.1) *and calculate* D_n *and/or* W_n^2 *accordingly.*
(D) *Define the wild bootstrap residual by* $\varepsilon^*_{\ell,i} = \hat{\varepsilon}_i \cdot \tau^*_{\ell,i}$, *where* $(\tau^*_{\ell,i})_{1\le\ell\le K, 1\le i\le n}$ *are i.i.d. Rademacher random variables which are independent of* $(Y_1, X_1), \ldots, (Y_n, X_n)$, *for* $i = 1, \ldots, n$ *and* $\ell = 1, \ldots, K$.
(E) *Set* $X^*_{\ell;i} = X_i$, *for* $i = 1, \ldots, n$ *and* $\ell = 1, \ldots, K$.
(F) *Set* $Y^*_{\ell;i} = m(\beta_n^\top X^*_{\ell;i}) + \varepsilon^*_{\ell;i}$, *for* $i = 1, \ldots, n$ *and* $\ell = 1, \ldots, K$.
(G) *Calculate the LSE* $\hat{\beta}^*_{\ell;n}$ *based on* $(Y^*_{\ell;1}, X^*_{\ell;1}), \ldots, (Y^*_{\ell;n}, X^*_{\ell;n})$, *for* $\ell = 1, \ldots, K$.
(H) *Obtain the MEP* $R^{*1}_{\ell;n}$ *according to* (6.2), $D^*_{\ell;n}$ *and/or* $W^{*2}_{\ell;n}$, *for* $\ell = 1, \ldots, K$.
(I) *Determine the* $p-$*value of* D_n *within the simulated* $D^*_{\ell;n}$, $1 \le \ell \le K$ *and/or the* $p-$*value of* W_n^2 *within the simulated* $W^{*2}_{\ell;n}$, $1 \le \ell \le K$, *respectively.*

For this section, we generate artificial data following the very simple model

$$Y = \sin(0.5X) + \varepsilon,$$

where X is uniformly distributed and ε is normally distributed.

```r
set.seed(123,kind ="Mersenne-Twister",normal.kind ="Inversion")
gen_data <- function(N = 200, sd = 0.2)  {
  dplyr::mutate(
    data.frame(X = runif(N, min = 6, max = 14)),
    mu = sin(0.5 * X),
    epsilon = rnorm(N, sd = sd),
    Y = mu + epsilon)
}
nonlinear <- gen_data()
```

Assuming that we did not know the model, Fig. 6.14 clearly indicates a polynomial relation.

```r
GGally::ggpairs(nonlinear[, c("X", "Y")])
```

One way to model such data (within a linear model) is to start with a simple model, in this case a polynomial of order two, and then gradually increase the complexity by increasing the degree of a polynomial. If there are some indications (maybe due

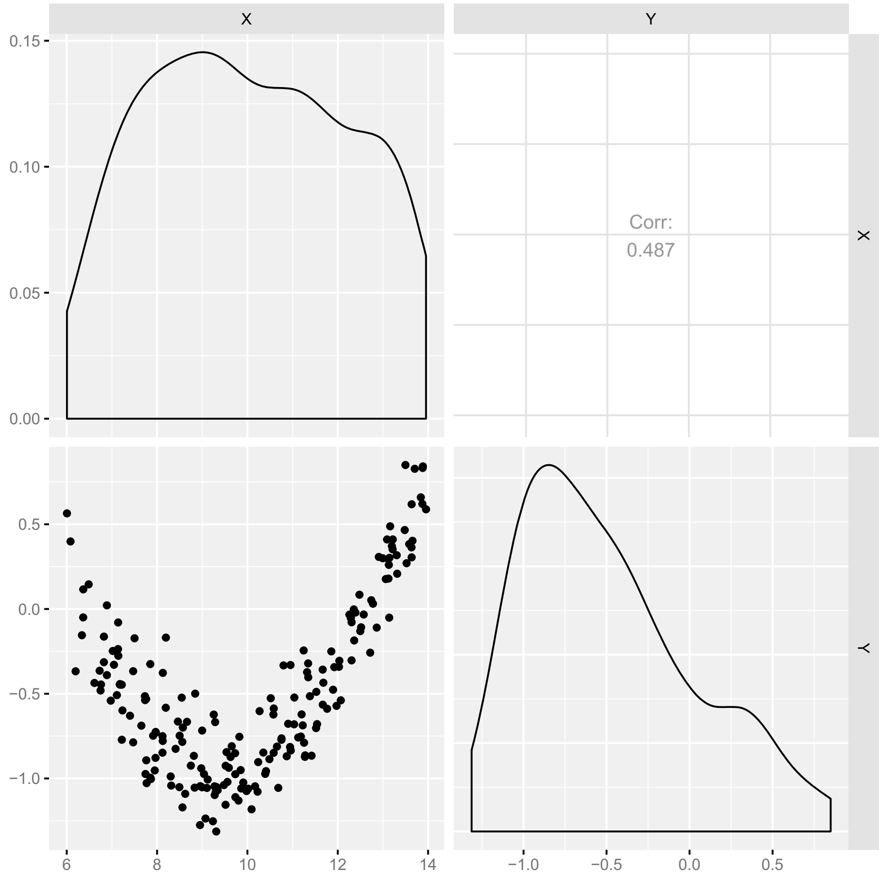

Fig. 6.14 Scatterplot of artificial data following $Y = \sin(0.5X) + \varepsilon$

to the theory of the problem at hand) that the pattern follows the sine function and one still wishes to model it with a linear model, it makes more sense to add the terms from the corresponding Taylor series, which in case of the sine function are only odd monomials. The common diagnostic plots for linear models do not indicate serious problems for a simple quadratic model, see Fig. 6.15

```
par(mfrow = c(2,2))
quadratic_fit <- glm(Y ~ X + I(X^2), data = nonlinear)
plot(quadratic_fit)
```

However, the wild bootstrap GOF test rejects the model (p-value = 0.024). The following two sections will implement the GOF test based on the wild bootstrap for the model specified in this section and apply it in a simulation study.

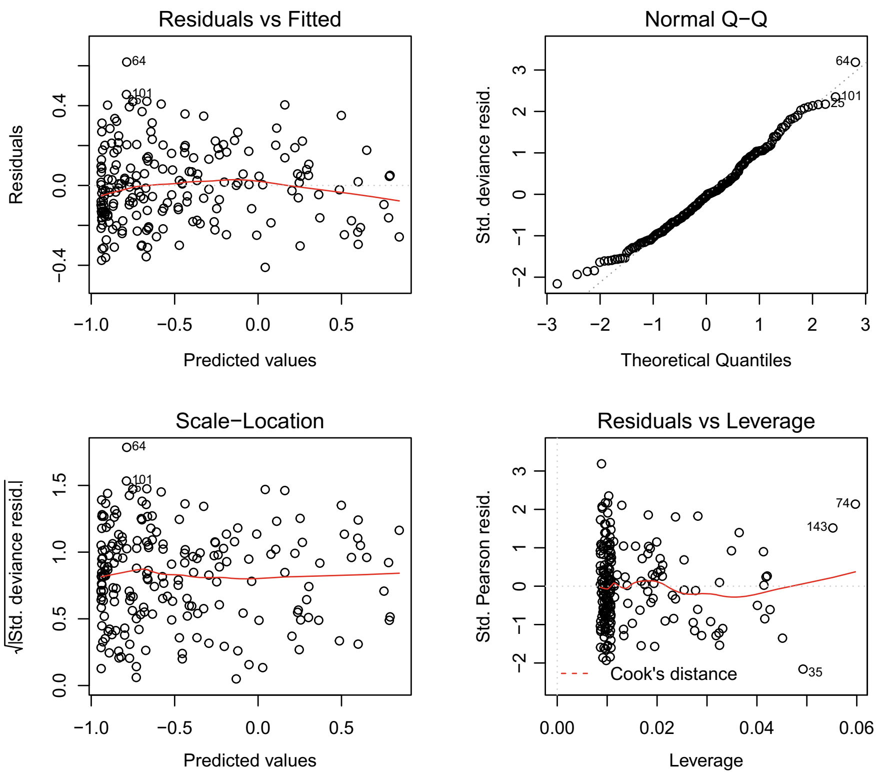

Fig. 6.15 Diagnostic plots of a linear model that incorporates the first two polynomial degrees, where $Y = \sin(0.5X) + \varepsilon$

6.2.1 Implementation

The implementation of the wild bootstrap is tailored to the model

$$Y = \sin(aX) + \varepsilon.$$

However, it is very similar to the implementation of GOF test using the parametric bootstrap. Basically, the difference is how the Y is generated and how the model is fitted. But the implementation is not so generic as for the parametric case. The calculation of $\beta_n^\top X$ is tailored to the situation that X is univariate. The main reason is that we will also use "minpack::nlsLM" for fitting a model and there seems to be no easy way to get this linear combination from the fit. Therefore, we prefer this simple implementation instead of a more generic but more complicated version.

```
rrademacher <- function(n) {
  2 * rbinom(n = n, size = 1, prob = 1/2) - 1
}

gof_wb_boot <- function(model, data, B = 1000) {
  # mod  - a model fit (stats::glm or minpack.lm::nlsLM)
  #   + for models based on nlsLM it is assumed that
  #     the formula is of the type fun(para * X), where
  #     'para' is the parameter and 'X' is the covariate
  # data - observed data, column with name 'X'
  #        is the only covariate
  # B    - number of bootstrapped MEPs

  # progress bar that appears if calculations will take more
  # than 1 second
  pb <- dplyr::progress_estimated(B, min_time = 1)

  if (inherits(model, "nls")) {
    # only models of type fun(para * X) are supported
    est_b_time_x <- data$X * coef(model)
  } else {
    est_b_time_x <- predict(model, type = "response")
  }

  # statistics for the original model
  Wn2 <- CvM(model, est_b_time_x = est_b_time_x)
  epsilon_hat <- residuals(model)
  y_hat <- predict(model)

  # copy to build up the bootstrap data set
  data_boot <- data
  y_name <- all.vars(formula(model), max.names = 1)

  # statistics for the boostrap models
  Wn2_boot <- sapply(seq_len(B), function(i) {

    pb$tick()$print() # print progress

    # according to Step E only the target/dependent
    # variable needs to be updated
    tau <- rrademacher(length(epsilon_hat))
    data_boot[[y_name]] <- y_hat + tau * epsilon_hat

    # refit the model using the bootstrapped data set
    m_boot <- update(model, formula. = formula(model),
                     data = data_boot)
    # statistic for the boostrapped model fit
    CvM(m_boot, est_b_time_x)
  })
  ret <- list(Wn2_boot = Wn2_boot,
              Wn2 = Wn2,
              pvalue_cvm = mean(Wn2_boot > Wn2))
  ret
}
```

6.2.2 Artificial Data

We use linear models with different polynomial degrees in this simulation and the wild bootstrap GOF test to check them. Note, in Sect. 6.3, we will come back to a similar situation and compare the parametric and wild bootstrap version of the GOF test.

```
gof_boot_nls <- function(data) {
 # performs a least square estimation
 # for sin(a * X) and a
 # semi-parametric GOF test
 # returns the corresponding p-values

 # data - original data set

  fit <- minpack.lm::nlsLM(Y ~ sin(a * X),
             data = data,
             start = c(a = 0.5),
             control = nls.control(maxiter = 500))
  gof <- gof_wb_boot(fit, data, B = 100)
  gof$pvalue_cvm
}

gof_boot_lm <- function(data, formula_str) {
 # fits a guassian model, performs
 # semi-parametric GOF test and returns
 # the corresponding p-values

 # data - original data set
 # formula_str - a formula as a string

  frml <- as.formula(formula_str)
  fit <- glm(frml, data = data, family = gaussian())
  gof <- gof_wb_boot(fit, data, B = 100)
  gof$pvalue_cvm
}
dg <- simTool::expand_tibble(proc = "gen_data", N = 100,
                             sd = 0.2)
pg <- dplyr::bind_rows(
  simTool::expand_tibble(fun = "gof_boot_nls"),
  simTool::expand_tibble(
    fun = "gof_boot_lm",
    formula_str = c("Y ~ X + I(X^2)",
                "Y ~ X + I(X^2) + I(X^3)",
                "Y ~ X + I(X^3) + I(X^5)"
                ))
)
eg <- simTool::eval_tibbles(
 data_grid = dg, proc_grid = pg,
 replications = 100, ncpus = 3,
 cluster_global_objects = ls())
```

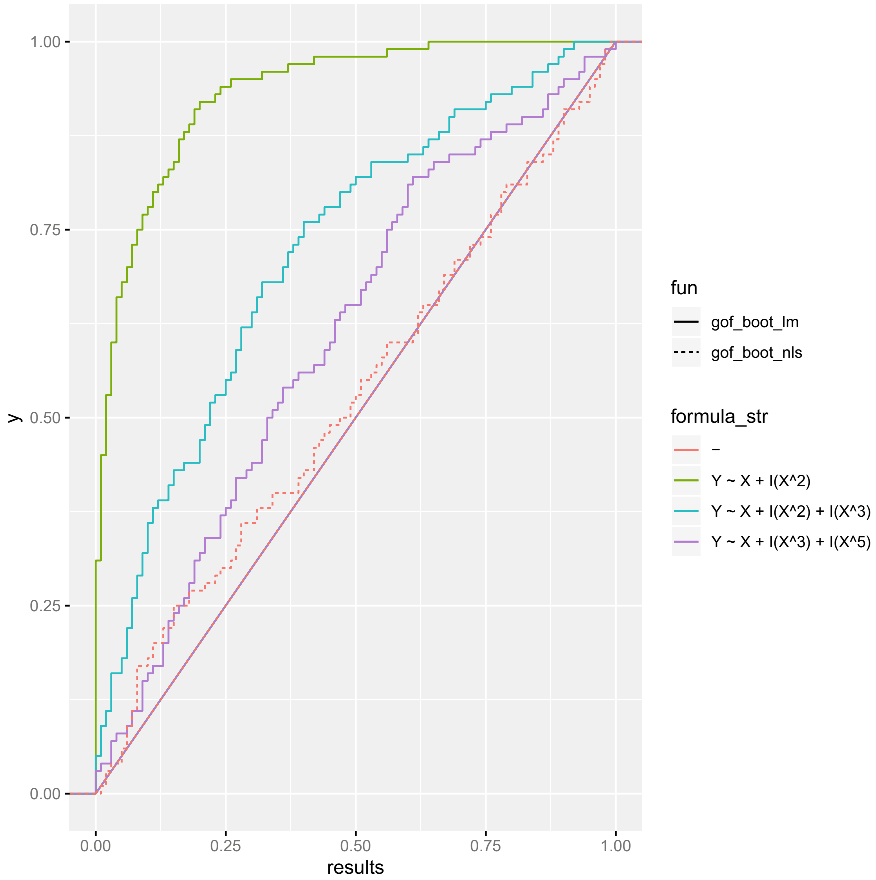

Fig. 6.16 Empirical cumulative distribution function of the p-values of GOF test based on the wild bootstrap for the semi-parametric model and of the parametric GOF test for the linear models, $Y = \sin(0.5X) + \varepsilon$ is conditionally normal distributed

Figure 6.16 shows the results of the simulation study. First note that the wild bootstrap shows a uniform distribution for the $p-$values, as expected since this is the correct model. Furthermore, the quadratic model as before shows a high chance to be rejected by the GOF test. Of course, increasing the complexity results in less rejections which is plausible because the sine function can be approximated this way. However, it is more effective to just use polynomials of an odd degree, which is also reflected by Fig. 6.16.

```
eg$simulation %>%
  dplyr::mutate(formula_str = ifelse(is.na(formula_str), "-",
                                     formula_str)) %>%
  ggplot(aes(x = results, linetype = fun, color = formula_str)) +
  stat_ecdf() +
  stat_function(fun = identity)
```

6.3 Comparison of the GOF Tests under the Parametric and Semi-parametric Setup

In general, we can assume that the parametric GOF test performs better than the corresponding wild bootstrap version, simply because in the parametric case we have more information and also explicitly use it. Here, we want to briefly compare both bootstrap versions. In order to do this, we use a similar setting as in Sect. 6.2. There we used the sine function to generate non-linear relation, i.e.,

$$\mathbb{E}(Y|X) = \sin(aX),$$

where X is uniformly distributed on the interval [6, 14] and the error followed a centered normal distribution. Here, we use two different distribution for Y, i.e., a normal distribution and a Poisson distribution. Since the Poisson distribution does not allow negative values, we extend the model by considering

$$\mathbb{E}(Y|X) = 4 + 2\sin(0.5X).$$

However, the corresponding Taylor series still contains only odd monomials and the simulations are also restricted to models of various polynomial degrees. So within the framework of generalized linear models, like in Sect. 5.3, all models we pick will be wrong. Within the semi-parametric setting we could choose the correct model but applying the GOF test has currently no theoretical foundation because $\beta_1 + \beta_2 \sin(\beta_3 X)$ cannot be written as $m(\beta^\top X)$. Figure 6.17 shows the results with conditionally normal distribution.

```
eg_para_vs_wb$simulation %>%
  ggplot(aes(x = results, linetype = fun, color = rhs)) +
  stat_ecdf() +
  facet_grid(N~., labeller = label_both) +
  stat_function(fun = identity, color = "black")
```

Since the least square estimator and maximum likelihood estimator are the same in this setting, any differences are only due to the generation of Y in the bootstrap world. At the first glance, it seems that wild bootstrap outperforms the parametric bootstrap for the small sample size $N = 10$. Although the theory is currently not rich enough, we applied also the semi-parametric GOF test, which shows that the GOF test is too liberate, i.e., the red dashed curve is above the diagonal. This, of course, is just an indicator of why the performance seems to be better. Furthermore, the models that are closer to the true model got rejected more often, which, of course, is a bit unusual. Another indicator that the performance advantage of the semi-parametric GOF test is probably spurious is that the same simulation based on sample size $N = 200$ shows that the performance of the semi-parametric GOF test degrades for the models with higher polynomials. For instance, according to ecdf of the $p-$values for the semi-parametric GOF test under the model $X + I(X^3) + I(X^5)$, around 25% of the $p-$values are below 0.05 for $N = 10$ and this decreases to around 5% for

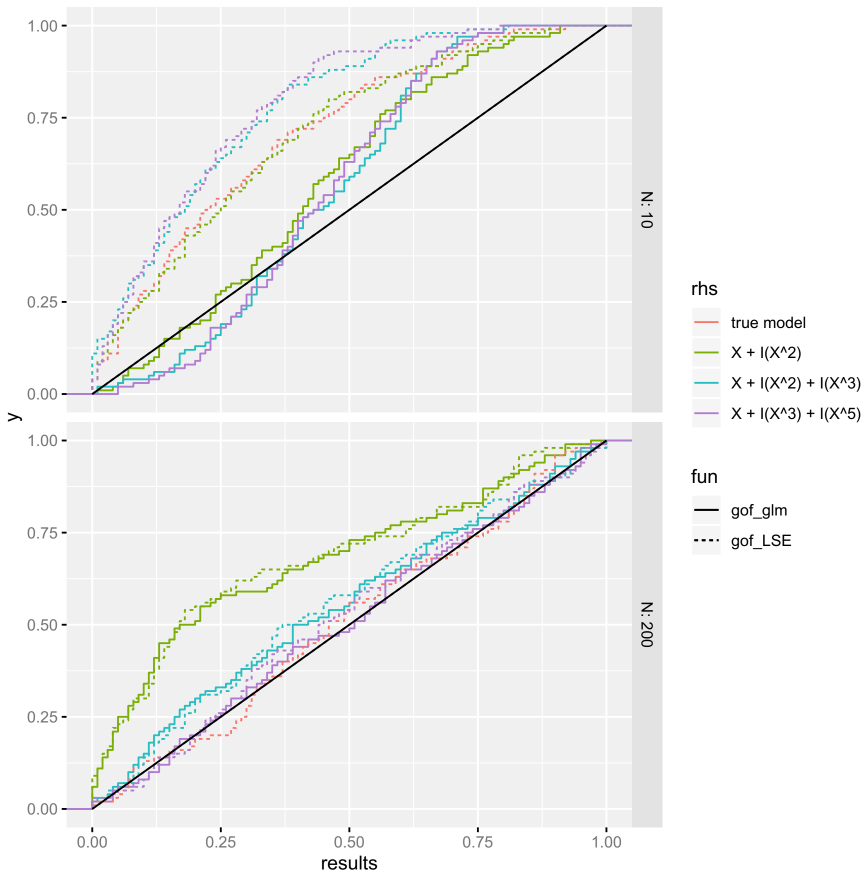

Fig. 6.17 Empirical cumulative distribution function of the $p-$values of GOF test based on the wild bootstrap (gof_LSE) for the semi-parametric model and of the parametric GOF test for the linear models (gof_glm), where $\mathbb{E}(Y|X) = 4 + 2\sin(0.5X)$ and Y is conditionally normal distributed. Sample size of the generated dataset is denoted by N. The right-hand side (rhs) describes which model was tested

$N = 200$. Note also that Fig. 6.17 shows that the $p-$values of the semi-parametric GOF test under the true model now seem to follow a uniform distribution. In this particular situation, both methods seem to have roughly equal performance, where for small sample size the semi-parametric GOF test may be too liberal. One reason why the semi-parametric GOF test is too liberal could be the Radermacher random variables, because this only changes the signs of the residuals and hence this only introduces little variation in the bootstrap datasets if the sample size is small.

Changing the conditional distribution to Poisson changes the results as expected, i.e., that the parametric GOF test results in more rejections if the sample size is

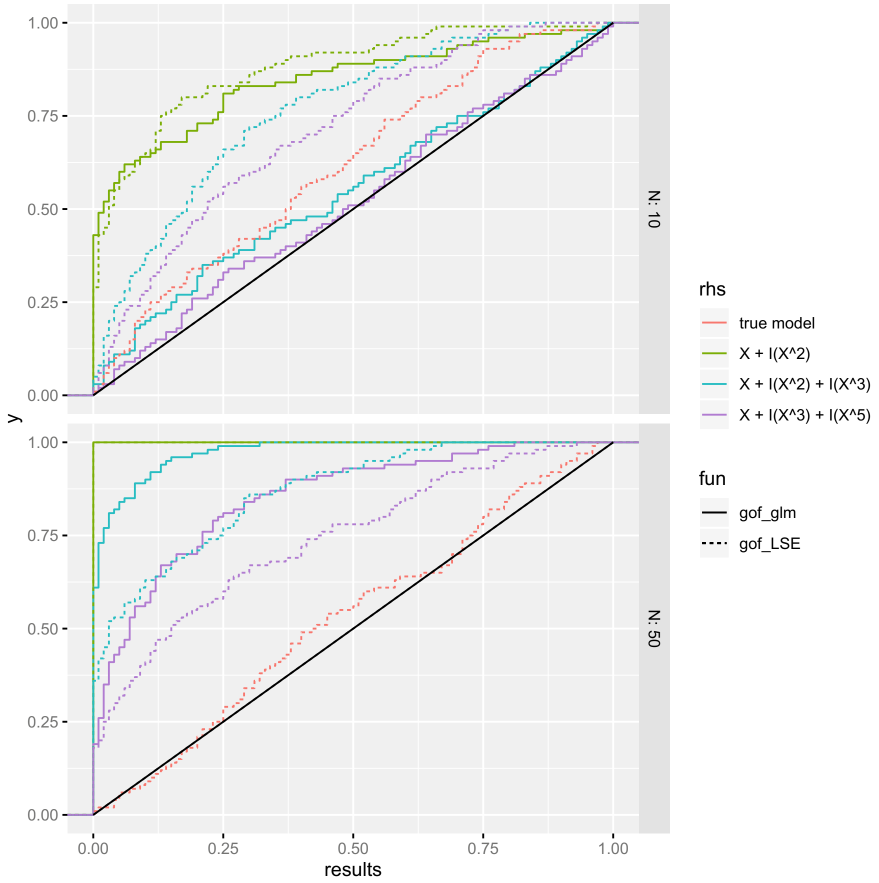

Fig. 6.18 Empirical cumulative distribution function of the p-values of GOF test based on the wild bootstrap (gof_LSE) for the semi-parametric model and of the parametric GOF test for the linear models (gof_glm), where $\mathbb{E}(Y|X) = 4 + 2\sin(0.5X)$ and Y is conditionally Poisson distributed. Sample size of the generated dataset is denoted by N. The right-hand side (rhs) describes which model was tested

sufficient. Furthermore, both GOF tests indicate that it is more efficient to use only odd polynomial degrees which corresponds also to our expectations, see Fig. 6.18.

```
eg_para_vs_wb2$simulation %>%
  ggplot(aes(x = results, linetype = fun, color = rhs)) +
  stat_ecdf() +
  facet_grid(N~., labeller = label_both) +
  stat_function(fun = identity, color = "black")
```

Again, for small sample size it seems that the performance gain of the semi-parametric GOF test is probably spurious for same reasons as before.

6.4 Mathematical Framework: Marked Empirical Processes

This excursion outlines some fundamental results of marked empirical processes (MEP) based on residuals given in Stute (1997) and Stute and Zhu (2002).

Since the following explanations are quite complex in their notation, we provide a short guideline in advance. In general, all analyzed MEPs are cumulative sum (cusum) error processes, which each propagate in a one-dimensional direction. The direction of the propagation is given within each indicator variable. If the indicator variable contains an estimated parameter, we speak here of an MEP that propagates in an estimated direction. We will consider three basic types of MEPs.

1. If the process is based on the true error and unfolds in a fixed, not estimated direction, then the process is called BMEP in the following and is always denoted with R_n.
2. If the MEP is based on estimated errors and unfolds in a fixed direction, then R_n^1 is chosen to denote the process. Processes of this kind are called EMEP in the following text.
3. Finally, there is a third type which will be called EMEPE. These processes are based on estimated errors and also on an estimated propagation direction. For these processes, we use the notation $\bar{R}_n^1$.

Furthermore, in our considerations, we make constant use of mathematical rules for conditional expectation without explicitly stating them in each case. A list of these rules can be found in Shorack (2000, Chapter 8.4 – 8.6). Only the concept of conditional variance is explained in more detail at this point. Let $Y \in L^2(\Omega, \mathscr{A}, \mathbb{P})$ and X be another random variable over $(\Omega, \mathscr{A}, \mathbb{P})$. Then we denote in the following with

$$\mathrm{VAR}(Y|X) = \mathbb{E}\big((Y - \mathbb{E}(Y|X))^2 \big| X\big)$$

the conditional variance of Y given X.

The space $D[0, 1]$ provided with the Skorokhod topology is the metric space for investigating the convergence in distribution of the empirical process, see Billingsley (1968, Chapter 3). The processes to be examined in this section will usually be in $D[-\infty, \infty]$.

Definition 6.3 Define $D[-\infty, \infty]$ as the collection of all right continuous functions $f : \mathbb{R} \to \mathbb{R}$ whose left-sided limits exist and for which $\lim_{x\to\infty} f(x)$ and $\lim_{x\to-\infty} f(x)$ exist in $\mathbb{R}$ also applies.

Remark 6.4 Now consider a continuous, strictly increasing transformation

$$A : \quad [-\infty, \infty] \longrightarrow [0, 1].$$

For example, a continuous, strictly increasing distribution function for A can be used here. Then the transformation

$$T : D[0, 1] \ni f \longrightarrow T(f) = f \circ A \equiv f_A \in D[-\infty, \infty]$$

is a bijective transformation. Let s denote the Skorokhod metric on $D[0, 1]$. Then

$$\begin{aligned} s_A : D[-\infty, \infty] \times D[-\infty, \infty] \ni (f_A, g_A) \longrightarrow & s_A(f_A, g_A) \\ & = s(f_A \circ A^{-1}, g_A \circ A^{-1}) \end{aligned}$$

defines a metric on $D[-\infty, \infty]$ which makes T to an isometric transformation. Thus, we can identify $D[-\infty, \infty]$ with $D[0, 1]$. Among other things, this isometry states that all theorems for convergence in distribution with respect to $D[0, 1]$ can now be transferred to $D[-\infty, \infty]$ accordingly. So we do not need to consider the limitation by the domain $[0, 1]$ anymore, if the process to be considered is in $D[-\infty, \infty]$.

6.4.1 The Basic MEP

Definition 6.5 Let $(Y_1, X_1), \ldots, (Y_n, X_n)$ be an i.i.d. sequence in $\mathbb{R}^2$ such that $\mathbb{E}(|Y|) < \infty$ and denote the conditional expectation of Y given $X = x$ by $m(x)$, that is,

$$\mathbb{E}(Y \mid X = x) = m(x).$$

Then

$$[-\infty, \infty] \ni x \longrightarrow R_n(x) := n^{-1/2} \sum_{i=1}^{n} \big(Y_i - m(X_i)\big) \mathrm{I}_{\{X_i \le x\}} \in \mathbb{R} \tag{6.3}$$

defines the *basic marked empirical process* (BMEP).

Note that $R_n(x)$ is defined for $x = \pm\infty$ by

$$R_n(-\infty) = 0 \quad \text{and} \quad R_n(\infty) = n^{-1/2} \sum_{i=1}^{n} \big(Y_i - m(X_i)\big).$$

This extends R_n continuously from $\mathbb{R}$ to $[-\infty, \infty]$ and allows to handle $\sup_{x\in\mathbb{R}} \big|R_n(x)\big|$, since $R_n \in D[-\infty, \infty]$.

Asymptotic analysis of the simple empirical process often uses the transformation of the process to the uniformly empirical process, see Sect. 3.4. A similar procedure is also possible with the BMEP, as described in Stute (1997) and as we will illustrate in detail now.

Since $F^{-1} \circ F(X) = X$ with probability 1, compare Shorack and Wellner (1986, Chapter 1, Proposition 3, Equation (27)), where F and F^{-1} denote the distribution and quantile function of X, respectively, we get with probability 1

$$\begin{aligned}
R_n(x) &= n^{-1/2} \sum_{i=1}^{n} \left(Y_i - m \circ F^{-1} \circ F(X_i)\right) \mathrm{I}_{\{F^{-1} \circ F(X_i) \le x\}} \\
&= n^{-1/2} \sum_{i=1}^{n} \left(Y_i - m \circ F^{-1}(F(X_i))\right) \mathrm{I}_{\{F(X_i) \le F(x)\}} \\
&= \hat{R}_n(F(x)),
\end{aligned}$$

where

$$\hat{R}_n(u) = n^{-1/2} \sum_{i=1}^{n} \left(Y_i - m \circ F^{-1}(F(X_i))\right) \mathrm{I}_{\{F(X_i) \le u\}}, \quad \text{for } 0 \le u \le 1. \tag{6.4}$$

While $R_n \in D[-\infty, \infty]$, we have $\hat{R}_n \in D[0, 1]$ and we can interpret $\hat{R}_n$ as a BMEP based on the $F(X)-$sample if $m \circ F^{-1}(u) = \mathbb{E}(Y \mid F(X) = u)$, for $\mathbb{P}_{F(X)}$ almost all $0 \le u \le 1$.

Lemma 6.6 *Let $Y \in L^2(\Omega, \mathscr{A}, \mathbb{P})$. Denote the distribution and quantile function of X by F and F^{-1}, respectively. With $m(x) = \mathbb{E}(Y \mid X = x)$ and $\sigma^2(x) = VAR(Y \mid X = x)$ we get*

(i) The conditional expectation of Y given $F(X) = u$, that is $\mathbb{E}(Y \mid F(X) = u)$, is well defined for $\mathbb{P}_{F(X)}$ almost all $0 \le u \le 1$ and

$$\mathbb{E}(Y \mid F(X) = u) = \mathbb{E}(Y \mid X = F^{-1}(u)) = m \circ F^{-1}(u).$$

(ii) The conditional variance of Y given $F(X) = u$, that is $VAR(Y \mid F(X) = u)$, is well defined for $\mathbb{P}_{F(X)}$ almost all $0 \le u \le 1$ and

$$VAR(Y \mid F(X) = u) = VAR(Y \mid X = F^{-1}(u)) = \sigma^2 \circ F^{-1}(u).$$

(iii) If F is continuous then $F(X)$ is uniformly distributed on $[0, 1]$ and, with $U = F(X)$, the last equalities read as follows:

$$\mathbb{E}(Y \mid U = u) = m(F^{-1}(u)), \quad VAR(Y \mid U = u) = \sigma^2(F^{-1}(u)),$$

for $\mathbb{P}_U$ almost all $0 \le u \le 1$.

Proof For the first equation, let B be an arbitrarily chosen Borel set of the unit interval. Then

$$\begin{aligned}
\int \mathrm{I}_{\{F(X) \in B\}} \, Y \, \mathrm{d}\mathbb{P} &= \int \mathrm{I}_{\{F(x) \in B\}} \, \mathbb{E}(Y \mid X = x) \, \mathbb{P}_X(\mathrm{d}x) \\
&= \int \mathrm{I}_{\{F(x) \in B\}} \, \mathbb{E}(Y \mid X = F^{-1} \circ F(x)) \, \mathbb{P}_X(\mathrm{d}x) \\
&= \int \mathrm{I}_{\{u \in B\}} \mathbb{E}(Y \mid X = F^{-1}(u)) \, \mathbb{P}_{F(X)}(\mathrm{d}u),
\end{aligned}$$

where the second equality is based on $\mathbb{P}_X(\{x \in \mathbb{R} : F^{-1}(F(x)) = x\}) = 1$, compare Shorack and Wellner (1986, Chapter 1, Proposition 3, Equation (27)). This proves, according to Shorack (2000, Chapter 8, Notation 4.1),

$$m(F^{-1}(u)) = \mathbb{E}(Y \mid F(X) = u), \quad \text{for } \mathbb{P}_{F(X)} \text{ almost all } u.$$

This proves (i). Analogously, part (ii) can be shown, and (iii) is obvious. □

Remark 6.7 The last lemma tells us that a change of the regressor from X to $F(X)$ causes a change of the regression function from m to $m \circ F^{-1}$, which is now the regression function of Y with respect to $F(X)$. Thus, we can transform the BMEP to $\hat{R}_n(F(x))$, where

$$\hat{R}_n(u) = n^{-1/2} \sum_{i=1}^{n} \big(Y_i - m \circ F^{-1}(F(X_i))\big)\, \mathrm{I}_{\{F(X_i) \le u\}}, \quad \text{for } 0 \le u \le 1,$$

which is now the BMEP corresponding to $(Y_1, F(X_1)), \ldots, (Y_n, F(X_n))$. In addition, if F is continuous then $F(X)$ is uniformly distributed on the unit interval and we can set $U_i = F(X_i)$. Now $\hat{R}_n$ is the transformed R_n to the uniform case. It is the counterpart to the uniform empirical process.

As we discussed in Sect. 4.4, the asymptotic behavior of the empirical process is the mathematical backbone in the context of model diagnostics for parametric distribution families of i.i.d. observations in $\mathbb{R}$. The main asymptotic result of the BMEP is given in the next theorem, compare Stute (1997, Theorem 1.1), and it shows that the BMEP has the potential to play the empirical processes counterpart in the regression context. However, in the following theorem, we will not make use of the transformation to the uniform case as it was done in the proof of Stute (1997, Theorem 1.1). The reasons for this are discussed after the proof at the end of this section. Instead, we will refer to Remark 6.4 in the proof.

Theorem 6.8 *Assume that* $\mathbb{E}(Y^2) < \infty$ *and*

$$H : [-\infty, \infty] \ni u \longrightarrow H(u) = \int I_{\{x \le u\}}\, \sigma^2(x)\, F(\mathrm{d}x) \in \mathbb{R} \tag{6.5}$$

is continuous. Then

$$R_n \longrightarrow R_\infty \quad \textit{in distribution in the space } D[-\infty, \infty].$$

R_∞ *is a centered Gaussian process with covariance function*

$$K(s,t) = H(s \wedge t) = \int I_{\{x \le s \wedge t\}} \sigma^2(x)\, F(\mathrm{d}x), \tag{6.6}$$

where F denotes the distribution function of X, $s \wedge t = \min(s, t)$, and $\sigma^2(x) = VAR(Y \mid X = x)$, the conditional variance of Y given $X = x$.

Remark 6.9 The covariance function of the limiting centered Gaussian process R_∞ is identical to the covariance function of the process $B(H)$, where B is a standard Brownian motion on the positive real line with H given under (6.5). Therefore, the paths of R_∞ are continuous.

Corollary 6.10 *If $\mathbb{E}(Y^2) < \infty$ and F is continuous, the continuity assumption (6.5) is fulfilled.*

Proof (of Theorem 6.8) The following proof is based on Stute (1997, Proof of Theorem 1.1) with some adjustments which are discussed in Example 6.12. It is an application of Billingsley (1968, Theorem 15.6).

Conditioning on $X_1, \ldots, X_n$ guarantees

$$\mathbb{E}(R_n(x)) = n^{-1/2} \sum_{i=1}^{n} \mathbb{E}\big(\mathrm{I}_{\{X_i \le x\}} \mathbb{E}\big((Y_i - m(X_i)) \big| X_i\big)\big) = 0,$$

for every $-\infty \le x \le \infty$. Since the terms in the sum of $R_n(x)$ are centered (expectation is 0) and i.i.d., we get in addition that

$$\mathbb{E}(R_n^2(x)) \le \mathbb{E}\big((Y - m(X))^2\big) \le \mathbb{E}(Y^2) < \infty.$$

Overall, $R_n(x) \in L_0^2(\Omega, \mathscr{A}, \mathbb{P})$, the space of square-integrable centered functions on $(\Omega, \mathscr{A}, \mathbb{P})$, for $-\infty \le x \le \infty$.

To apply Billingsley (1968, Theorem 15.6), we will show that the finite-dimensional distributions (fidis) of R_n converge to those of R_∞. Take $-\infty \le x_1, \ldots, x_k \le \infty$, for $k \in \mathbb{N}$ and apply the multivariate central limit theorem, Billingsley (1995, Theorem 29.5) to get that

$$(R_n(x_1), \ldots, R_n(x_k)) \longrightarrow \mathscr{N}(0, \Sigma), \quad \text{as } n \to \infty,$$

in distribution, where $\Sigma = (\sigma_{i,j})_{1 \le i,j \le k}$ is the covariance matrix defined by

$$\sigma_{i,j} = \mathrm{COV}(R_n(x_i), R_n(x_j)), \quad \text{for } 1 \le i, j \le k.$$

Since

$$\begin{aligned}\mathrm{COV}(R_n(x_i), R_n(x_j)) &= \mathbb{E}\big(\mathrm{I}_{\{X \le x_i\}}\, \mathrm{I}_{\{X \le x_j\}}\, (Y - m(X))^2\big) \\ &= \int \mathrm{I}_{\{t \le x_i \wedge x_j\}} \sigma^2(t)\, F(\mathrm{d}t) = K(x_i, x_j),\end{aligned}$$

the first part of the proof is done.

To prove tightness, we adapt Billingsley (1968, (15.21) in Theorem 15.6) according to Remark 6.4 to $D[-\infty, \infty]$.

For $-\infty \le x_1 \le x \le x_2 \le \infty$, set

$$\alpha_i = \mathrm{I}_{\{x_1 < X_i \le x\}}\big(Y_i - m(X_i)\big)$$

and

$$\beta_i = \mathrm{I}_{\{x < X_i \le x_2\}}\big(Y_i - m(X_i)\big)$$

to obtain

$$\mathbb{E}\Big((R_n(x) - R_n(x_1))^2 (R_n(x_2) - R_n(x))^2\Big) = n^{-2}\mathbb{E}\Big(\big(\sum_{1\le i\le n} \alpha_i\big)^2 \big(\sum_{1\le j\le n} \beta_j\big)^2\Big).$$

Due to the indicator functions involved in the definition of α_i and β_i, $\alpha_i\,\beta_i = 0$. Conditioning on X_i shows that $\mathbb{E}(\alpha_i) = 0 = \mathbb{E}(\beta_i)$, for $1 \le i \le n$. Furthermore, $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_n$ are i.i.d. sequences, where, in addition, α_i is independent from β_j, for $1 \le i \ne j \le n$. Overall, this results in

$$\begin{aligned} n^{-2}\mathbb{E}\Big(\big(\sum_{1\le i\le n} \alpha_i\big)^2 \big(\sum_{1\le j\le n} \beta_j\big)^2\Big) &= \frac{n-1}{n}\mathbb{E}(\alpha_1^2)\mathbb{E}(\beta_1^2) \\ &\le \big(H(x) - H(x_1)\big)\big(H(x_2) - H(x)\big) \\ &\le \big(H(x_2) - H(x_1)\big)^2. \end{aligned}$$

Since H is a nondecreasing, continuous function, the proof is complete. □

Remark 6.11 The continuity assumption (6.5) in Theorem 6.8 is not dispensable even though it does not appear in Stute (1997, Theorem 1.1). In the proof of Stute (1997, Theorem 1.1), it is noted for the verification of tightness that $\mathbb{E}(Y \mid U = u) = m(F^{-1}(u))$. This then implies the continuity of H, our assumption (6.5). However, the following example shows that $\mathbb{E}(Y \mid U = u) = m(F^{-1}(u))$ does not generally have to be true if F is discontinuous. Nevertheless, in the main application of the theorem continuity of F has to be guaranteed anyway and the missing assumption in Stute (1997, Theorem 1.1) does not affect its importance in statistical application!

Example 6.12 Let U be a uniformly distributed random variable defined on some probability space $(\Omega, \mathscr{A}, \mathbb{P})$. Set $X = \mathrm{I}_{\{U \le 0.5\}}$ and $Y = U$. The Bernoulli-distributed random variable X has distribution, respectively, quantile function

$$F(x) = \mathbb{P}(X \le x) = \begin{cases} 0 & : x < 0 \\ 0.5 & : 0 \le x < 1 \\ 1 & : x \ge 1 \end{cases}$$

$$F^{-1}(u) = \begin{cases} 0 & : 0 \le u \le 0.5 \\ 1 & : 0.5 < u \le 1 \end{cases},$$

respectively. Elementary calculation of $\mathbb{E}(Y \mid X = x)$, for $x \in \{0, 1\}$, yields

$$\mathbb{E}(Y \mid X = F^{-1}(u)) = m \circ F^{-1}(u) = 3/4 - F^{-1}(u)\big/2.$$

Since $\mathbb{E}(Y \mid U = u) = u$, we finally get

$$\mathbb{E}(Y \mid U = u) \neq m \circ F^{-1}(u), \quad \text{for all } 0 \leq u \leq 1.$$

This contradicts

$$\mathbb{E}(Y \mid U = u) = m \circ F^{-1}(u)$$

asserted in Stute (1997, Proof of Theorem 1.1). □

6.4.2 *The MEP with Estimated Model Parameters Propagating in a Fixed Direction*

The result obtained under Theorem 6.8 for the BMEP represents an initial theoretical basis which, however, still has to be extended for statistical applications. If, for example, there is a parameterized regression, then the corresponding parameter must be estimated and instead of the true regression function m we now consider a estimated regression function. If the true m is replaced by the estimated one in the BMEP, then the true errors are replaced by the estimated errors, i.e., by the residuals. This, of course, affects the limit distribution. The extension of the BMEP with estimated parameters goes back to Stute (1997) and we present the result here in our context.

Definition 6.13 Let $(Y_1, X_1), \ldots, (Y_n, X_n)$ be an i.i.d. sequence in $\mathbb{R}^2$ such that $\mathbb{E}(|Y|) < \infty$ and denote the conditional expectation of Y given $X = x$ by $m(x)$. Assume that m belongs to a parametric family

$$\mathscr{M} = \big\{ m(\cdot, \theta) \ : \ \theta \in \Theta \big\}$$

such that $m(x) = m(x, \theta_0)$, for some true parameter $\theta_0 \in \Theta \subset \mathbb{R}^p$. Let θ_n be an estimator of θ_0. Then

$$[-\infty, \infty] \ni x \longrightarrow R_n^1(x) := n^{-1/2} \sum_{i=1}^{n} \big(Y_i - m(X_i, \theta_n)\big) \, \mathrm{I}_{\{X_i \leq x\}} \in \mathbb{R} \tag{6.7}$$

defines the *estimated marked empirical process* (EMEP).

The direction of propagation of EMEP, which is determined by the indicator, is given here by $\mathbb{R}$, as with the BMEP itself, since the covariate X is real. Therefore, the process propagates in a fixed direction.

For the analysis of the asymptotic behavior of R_n^1, we now, of course, need further assumptions which we list below:

(i) $n^{1/2}(\theta_n - \theta_0) = n^{-1/2} \sum_{i=1}^n l(X_i, Y_i, \theta_0) + o_{\mathbb{P}}(1)$.
(ii) $\mathbb{E}(l(X_i, Y_i, \theta_0)) = 0$.
(iii) $L(\theta_0) = \mathbb{E}(l(X, Y, \theta_0) l^\top (X, Y, \theta_0))$ exists.
(iv) $w(x, \theta) = \partial m(x, \theta)/\partial\theta = (w_1(x, \theta), \ldots, w_p(x, \theta))^\top$ exists for θ in a neighborhood $V \subset \Theta$ of θ_0 and is continuous with respect to θ.
(v) There exists an F-integrable function $M(x)$ such that $|w_i(x, \theta)| \leq M(x)$ for all $\theta \in V \subset \Theta$, $1 \leq i \leq p$, and V given in 6.4.2(iv).

The first three conditions 6.4.2(i)–(iii) are usually met for least squares or maximum likelihood estimates.

Let $W(t, \theta) = (W_1(t, \theta), \ldots, W_p(t, \theta))^\top$ be defined by

$$W_i(t, \theta) = \mathbb{E}\big(w_i(X, \theta) \mathrm{I}_{\{X \leq t\}}\big).$$

Lemma 6.14 *Let θ_n converge in probability to θ_0 and assume that $\hat{\theta}_n : \mathbb{R} \to \Theta$ is a measurable function such that $\hat{\theta}_n(x)$ lies for each $x \in \mathbb{R}$ and $n \in \mathbb{N}$ on the line segment that connects θ_n and θ_0. If 6.4.2(iv) and (v) hold, then, for $1 \leq i \leq p$,*

(i) $\sup_{-\infty \leq t \leq \infty} \big| \int_{-\infty}^t w_i(x, \hat{\theta}_n(x)) - w_i(x, \theta_0) F_n(\mathrm{d}x) \big| = o_{\mathbb{P}}(1)$,
(ii) $\limsup_{n\to\infty} \sup_{-\infty \leq t \leq \infty} \big| \int_{-\infty}^t w_i(x, \theta_0)\, (F_n - F)(\mathrm{d}x) \big| = 0$ *w.p.1.*

Proof For $\varepsilon, \delta > 0$ we get by the Markov theorem,

$$\mathbb{P}\left(\sup_{-\infty \leq t \leq \infty} \left| \int_{-\infty}^t w_i(x, \hat{\theta}_n(x)) - w_i(x, \theta_0)\, F_n(\mathrm{d}x) \right| > \varepsilon \right)$$
$$\leq \frac{1}{\varepsilon} \mathbb{E}\left(\sup_{|\theta - \theta_0| < \delta} |w_i(X, \theta) - w_i(X, \theta_0)| \right) + \mathbb{P}\big(|\theta_n - \theta_0| > \delta\big),$$

for $1 \leq i \leq p$. Due to 6.4.2(v), the expectation on the right side is finite and the integrand converges to 0 according to 6.4.2(iv) if δ tends to 0. The first assertion now follows from the dominated convergence theorem and the assumed convergence of θ_n to θ_0.

For the second assertion, we fix $K > 0$ and apply 6.4.2(v) to obtain

$$\limsup_{n\to\infty} \sup_{-\infty \leq t \leq \infty} \Big| \int_{-\infty}^t w_i(x, \theta_0)\, (F_n - F)(\mathrm{d}x) \Big|$$
$$\leq \limsup_{n\to\infty} \sup_{|t| \leq K} \Big| \int_{-K}^t w_i(x, \theta_0)\, (F_n - F)(\mathrm{d}x) \Big|$$
$$+ \limsup_{n\to\infty} \int M(x) \mathrm{I}_{\{|x| > K\}}\, F_n(\mathrm{d}x) + \int M(x) \mathrm{I}_{\{|x| > K\}}\, F(\mathrm{d}x).$$

According to Theorem 5.66, the first term on the right side is identical to 0 for each fixed K w.p.1. Due to the SLLN, the second term is identical to the third term w.p.1. However, since the third term converges with $K \to \infty$ against 0, the second assertion is proven. □

The main result of this section is the following theorem, compare Stute (1997, Theorem 1.2).

Theorem 6.15 *Assume $\mathbb{E}(Y^2) < \infty$, F is continuous and conditions 6.4.2(i)–6.4.2 (v) are met. Then, under $m(\cdot) = m(\cdot, \theta_0)$, we have, uniformly in x,*

$$R_n^1(x) = R_n(x) - n^{1/2} \sum_{i=1}^{n} W^\top(x, \theta_0) l(X_i, Y_i, \theta_0) + o_{\mathbb{P}}(1), \quad \textit{as } n \to \infty.$$

Furthermore, R_n^1 converges in $D[-\infty, \infty]$ to a centered Gaussian process R_∞^1 with covariance function

$$\begin{aligned} K^1(s,t) = K(s,t) &+ W^\top(s, \theta_0) L(\theta_0) W(t, \theta_0) \\ &- W^\top(s, \theta_0) \mathbb{E}(I_{\{X \le t\}}(Y - m(X, \theta_0)) l(X, Y, \theta_0)) \\ &- W^\top(t, \theta_0) \mathbb{E}(I_{\{X \le s\}}(Y - m(X, \theta_0)) l(X, Y, \theta_0)). \end{aligned}$$

Proof By definition,

$$\begin{aligned} R_n^1(x) &= n^{-1/2} \sum_{i=1}^{n} \mathrm{I}_{\{X_i \le x\}} [Y_i - m(X_i, \theta_n)] \\ &= R_n(x) - n^{-1/2} \sum_{i=1}^{n} \mathrm{I}_{\{X_i \le x\}} [m(X_i, \theta_n) - m(X_i, \theta_0)]. \end{aligned}$$

A Taylor expansion of the terms of the sum results in

$$m(X_i, \theta_n) - m(X_i, \theta_0) = (\theta_n - \theta_0)^\top w(X_i, \theta_{ni}),$$

where θ_{ni} is between θ_n and θ_0. Hence,

$$\begin{aligned} R_n^1(x) &= R_n(x) - n^{1/2} (\theta_n - \theta_0)^\top n^{-1} \sum_{i=1}^{n} \mathrm{I}_{\{X_i \le x\}} w(X_i, \theta_{ni}) \\ &= R_n(x) - n^{1/2} (\theta_n - \theta_0)^\top n^{-1} \sum_{i=1}^{n} \mathrm{I}_{\{X_i \le x\}} (w(X_i, \theta_{ni}) - w(X_i, \theta_0)) \\ &\quad - n^{1/2} (\theta_n - \theta_0)^\top \left(n^{-1} \sum_{i=1}^{n} \mathrm{I}_{\{X_i \le x\}} w(X_i, \theta_0) - W(x, \theta_0) \right) \\ &\quad - n^{1/2} (\theta_n - \theta_0)^\top W(x, \theta_0) \\ &= R_n(x) - n^{1/2} (\theta_n - \theta_0)^\top W(x, \theta_0) + o_{\mathbb{P}}(1), \end{aligned}$$

uniformly in x, where the last equality follows from Lemma 6.14. Since $W(\cdot, \theta_0)$ is bounded, assumption 6.4.2(i) yields

$$n^{1/2}(\theta_n - \theta_0)^\top W(x, \theta_0) = n^{-1/2} \sum_{i=1}^{n} l^\top(X_i, Y_i, \theta_0) W(x, \theta_0) + o_{\mathbb{P}}(1), \tag{6.8}$$

uniformly in x which shows the first assertion of the theorem.

This representation of R_n^1 shows that the two sequences $(R_n^1)_{n\in\mathbb{N}}$ and $(\hat{R}_n^1)_{n\in\mathbb{N}}$, where

$$\hat{R}_n^1(x) = n^{-1/2} \sum_{i=1}^{n} \left((Y_i - m(X_i, \theta_0)) \, \mathrm{I}_{\{X_i \le x\}} - l^\top(Y_i, X_i, \theta_0) W(x, \theta_0) \right),$$

are asymptotically equivalent in the sense of Billingsley (1968, Theorem 4.1). Therefore, we can do the remaining part of the proof with $(\hat{R}_n^1)_{n\in\mathbb{N}}$.

In order to prove tightness of $(\hat{R}_n^1)_{n\in\mathbb{N}}$, it remains to show, by virtue of Theorem 6.8, that the right-hand side of (6.8) is also tight in $D[-\infty, \infty]$. By assumption 6.4.2(ii) and 6.4.2(iii) the sequence $(S_n)_{n\in\mathbb{N}}$, where $S_n = n^{-1/2} \sum_{i=1}^n l(X_i, Y_i, \theta_0)$, tends to a multivariate normal distribution and therefore is tight in $\mathbb{R}^p$. Since $W(\cdot, \theta_0)$ is a bounded deterministic continuous function, the sequence $(S_n^\top W(\cdot, \theta_0))_{n\in\mathbb{N}})$ is tight in $C[-\infty, \infty]$. Since $C-$tightness implies $D-$tightness, we have shown that $(\hat{R}_n^1)_{n\in\mathbb{N}}$ is tight in $D[-\infty, \infty]$.

The convergence of the fidis of $(\hat{R}_n^1)_{n\in\mathbb{N}}$ is a consequence of the multivariate CLT. Hence, it remains to calculate the covariance function $K^1(s, t)$ of R_∞^1 which is identical to the covariance function of the centered process $\hat{R}_n^1$. Recall the definition of $L(\theta_0)$ given under 6.4.2(iii) to get

$$\begin{aligned} \mathrm{COV}(\hat{R}_n^1(s), \hat{R}_n^1(t)) = {} & \mathbb{E}\left(\mathrm{I}_{\{X \le s \wedge t\}} (Y - m(X, \theta_0))^2\right) + W^\top(s, \theta_0) L(\theta_0) W(t, \theta_0) \\ & - W^\top(t) \mathbb{E}\left(\mathrm{I}_{\{X \le s\}} (Y - m(X, \theta_0)) l(Y, X, \theta_0)\right) \\ & - W^\top(s) \mathbb{E}\left(\mathrm{I}_{\{X \le t\}} (Y - m(X, \theta_0)) l(Y, X, \theta_0)\right). \end{aligned}$$

Note that the first term on the right side is the covariance of the BMEP limit process R_∞. □

Remark 6.16 Under the assumptions of Theorem 6.15, Shorack (2000, Chapter 12, Theorem 2.1 (1)) can be directly verified from the covariance function of R_∞^1. This shows that the limiting process R_∞^1 can be realized in $C[-\infty, \infty]$.

6.4.3 The MEP with Estimated Model Parameters Propagating in an Estimated Direction

Previously, all MEPs were based on one-dimensional random variables X, which were used to define the corresponding processes via the indicators $\mathrm{I}_{\{X\leq x\}}$. If the input variables X are multidimensional, then the indicator set $(-\infty, x]$ can be replaced by the quadrant with upper right corner x, i.e., by the $\{z \in \mathbb{R}^p : z_i \leq x_i, \text{ for } 1 \leq i \leq p\}$. However, if the model under consideration is a linear or a generalized linear model, then the multidimensional vector X acts on Y by a corresponding linear combination of its components and we can switch to a one-dimensional input, namely, to this linear combination. The corresponding process thus realizes itself again in $D[-\infty, \infty]$. However, we pay a price for this traceability to the one-dimensional case; the direction in which the process evolves is determined by the linear combination. It is thus based on the estimated parameters and therefore propagates in an estimated direction.

Definition 6.17 Let $(Y_1, X_1), \ldots, (Y_n, X_n)$ be an i.i.d. sequence in $\mathbb{R}^{1+p}$ such that $\mathbb{E}(|Y|) < \infty$ and denote the conditional expectation of Y given $X = x$ by $m(x)$. Assume that m belongs to a parametric family

$$\mathscr{M} = \left\{m(\cdot, \vartheta) : \vartheta = (\beta, \theta) \in \mathbb{R}^p \times \Theta \subset \mathbb{R}^{p+q}\right\}$$

such that $m(x) \equiv m(x, \vartheta_0) = m_0(\beta_0^\top x, \theta_0)$, for some true parameter $(\beta_0, \theta_0) \equiv \vartheta_0 \in \mathbb{R}^p \times \Theta$ and a Borel-measurable function $m_0 : \mathbb{R} \to \mathbb{R}$.

Let $\vartheta_n = (\beta_n, \theta_n)$ be an estimator of ϑ_0. Then

$$[-\infty, \infty] \ni u \longrightarrow \bar{R}_n^1(u) := n^{-1/2} \sum_{i=1}^{n} \left(Y_i - m_0(\beta_n^\top X_i, \theta_n)\right) \mathrm{I}_{\{\beta_n^\top X_i \leq u\}} \in \mathbb{R} \quad (6.9)$$

defines the *estimated marked empirical process in estimated direction* (EMEPE).

Remark 6.18 The specific form $m(x) = m_0(\beta_0^\top x, \theta_0)$ implies that w.p.1

$$\mathbb{E}(Y \mid X) = m(X) = m_0(\beta_0^\top X, \theta_0) = \mathbb{E}(Y \mid \beta_0^\top X), \quad (6.10)$$

i.e., $\mathbb{E}(Y \mid X)$ is measurable with respect to the smaller σ-field $(\beta_0^\top X)^{-1}(\mathscr{B}^*)$, where $\mathscr{B}^*$ denotes the Borel $\sigma-$field on $\mathbb{R}$. Be aware that this is a very restrictive condition, because it means that all the information from X concerning $\mathbb{E}(Y|X)$ is already stored in the information given by the projection of X onto the line defined by β_0! Note that in the following we will not distinguish between m and m_0, but will always use m, even if m_0 is meant. So instead of writing $m_0(\beta^\top x)$ we will use $m(\beta^\top x)$. Thus, the EMEPE will be

$$[-\infty, \infty] \ni u \longrightarrow \bar{R}_n^1(u) := n^{-1/2} \sum_{i=1}^{n} \left(Y_i - m(\beta_n^\top X_i, \theta_n)\right) \mathrm{I}_{\{\beta_n^\top X_i \leq u\}} \in \mathbb{R}. \quad (6.11)$$

Remark 6.19 If $m(x) = m(\beta_0^\top x, \theta_0)$ applies, then

$$e_i = Y_i - m(\beta_0^\top X_i, \theta_0)$$

denotes the true error. If ϑ_n is an estimator of ϑ_0, then

$$e_i(\vartheta_n) = Y_i - m(\beta_n^\top X_i, \theta_n)$$

defines the estimated error, that is, the residual. Furthermore, according to (6.10),

$$\mathbb{E}(e \mid X) = 0 = \mathbb{E}(e \mid \beta_0^\top X, \theta_0). \tag{6.12}$$

For the conditional variance with respect to $X = x$, we set

$$\sigma^2(x) = \mathbb{E}\big((Y - m(X))^2 \,\big|\, X = x\big) = \mathrm{VAR}\big(Y \mid X = x\big) \tag{6.13}$$

and define

$$\sigma^2_{\vartheta_0}(t) = \mathbb{E}\big((Y - m(\beta_0^\top X, \theta_0))^2 \,\big|\, \beta_0^\top X = t\big) = \mathrm{VAR}_{\vartheta_0}\big(Y \mid \beta_0^\top X = t\big). \tag{6.14}$$

For the functional limit theorem of the EMEPE, we again need some assumptions, which are directly derived from those of Sect. 6.4.2. We also need an additional assumption to control the estimated direction of EMEPE.

(i) $n^{1/2}(\vartheta_n - \vartheta_0) = n^{-1/2} \sum_{i=1}^n l(X_i, Y_i, \vartheta_0) + o_{\mathbb{P}}(1)$.
(ii) $\mathbb{E}(l(X_i, Y_i, \vartheta_0)) = 0$.
(iii) $L(\vartheta_0) = \mathbb{E}(l(X, Y, \vartheta_0) l^\top(X, Y, \vartheta_0))$ exists and is positive definite.
(iv) $w(x, \vartheta) = \partial m(x, \vartheta)/\partial \vartheta = (w_1(x, \vartheta), \ldots, w_{p+q}(x, \vartheta))^\top$ exists for ϑ in a neighborhood $V \subset \mathbb{R}^p \times \Theta$ of ϑ_0 and is continuous with respect to ϑ.
(v) There exists an $\mathbb{P}_X$-integrable function $M(x)$ such that $|w_i(x, \vartheta)| \leq M(x)$ for all $\vartheta \in V \subset \mathbb{R}^p \times \Theta$, $1 \leq i \leq p + q$, and V given in 6.4.3(iv).
(vi) The function

$$H: \quad \mathbb{R}^{p+1} \ni (\beta, u) \longrightarrow H(u, \beta) := \int \mathrm{I}_{\{\beta^\top X \leq u\}} \sigma^2(X)\, \mathrm{d}\mathbb{P} \in \mathbb{R}$$

is uniformly continuous in u at β_0.

Set $W(t) = W(t, \vartheta_0) = (W_1(t, \vartheta_0), \ldots, W_{p+q}(t, \vartheta_0))^\top$, where

$$W_i(t) = \mathbb{E}\left(w_i(X, \vartheta_0) \mathrm{I}_{\{\beta_0^\top X \leq t\}}\right). \tag{6.15}$$

The following technical lemma is of decisive importance for the functional limit theorem of the EMEPE that will follow later.

Lemma 6.20 *Assume that* $\mathbb{E}(Y^2) < \infty$, *6.4.3(vi), and* $m(x) = m(\beta_0^\top x, \theta_0)$ *holds for all* $x \in \mathbb{R}^p$. *Then we get for every* $\varepsilon > 0$*:*

$$\mathbb{P}\left(\sup_{u\in\mathbb{R}} \sup_{\{\beta : |\beta-\beta_0|\le\delta\}} \left|n^{-1/2}\sum_{i=1}^{n}\left(I_{\{\beta^\top X_i\le u\}} - I_{\{\beta_0^\top X_i\le u\}}\right)e_i\right| \ge \varepsilon\right) \longrightarrow 0,$$

as $\delta \to 0$.

Proof The proof is based on the theory of generalized empirical processes, as presented in the textbooks of van der Vaart and Wellner (1996) and Kosorok (2008), respectively.

For each $\beta \in \mathbb{R}^p$ and $u \in \mathbb{R}$ the set $H = H_{\beta,u} = \{x \in \mathbb{R}^p : \beta^\top x \le u\}$ defines a half-space in $\mathbb{R}^p$. Denote the set of all half-spaces of $\mathbb{R}^p$ by

$$\mathscr{H} = \left\{H = H_{\beta,u},\ \beta \in \mathbb{R}^p, u \in \mathbb{R}\right\}.$$

Based on this collection of sets we now define a function class of indicators through

$$\mathscr{F} = \left\{\mathrm{I}_{\{H\times\mathbb{R}\}} : H \in \mathscr{H}\right\}$$

and modify this class by multiplying the individual indicators by the function

$$h : \mathbb{R}^{p+1} \ni (x, y) \longrightarrow h(x, y) = y - m(\beta_0^\top x, \theta_0) \in \mathbb{R}$$

to get

$$\mathscr{F}_h = \left\{h\ \mathrm{I}_{\{H\times\mathbb{R}\}} : H \in \mathscr{H}\right\}.$$

Based on this collection of measurable functions, we consider the generalized empirical process $(\alpha_n(f))_{f\in\mathscr{F}_h}$,

$$\alpha_n(f) = n^{-1/2}\sum_{i=1}^{n} f(X_i, Y_i).$$

Note that according to (6.12) $\mathbb{E}(\alpha_n(f)) = 0$.

The paths of this generalized empirical process are elements of the space $l_\infty(\mathscr{F}_h)$, that is, the space of all function $l : \mathscr{F}_h \ni f \longrightarrow l(f) \in \mathbb{R}$ such that $\sup_{f\in\mathscr{F}_h} |l(f)| \equiv \|l\|_{\mathscr{F}_h} < \infty$. The metric $d_\infty(l_1, l_2) = \|l_1 - l_2\|_{\mathscr{F}_h}$ turns $l_\infty(\mathscr{F}_h)$ into the metric space $(l_\infty(\mathscr{F}_h), d_\infty)$.

First, we note that for every $f \in \mathscr{F}_h$ we have $f(X, Y) \in L^2(\Omega, \mathscr{A}, \mathbb{P})$. Furthermore, for each $f \in \mathscr{F}_h$ and $(x, y) \in \mathbb{R}^{p+1}$, $|f(x, y)| \le |h(x, y)|$, and $h(X, Y) \in L^2(\Omega, \mathscr{A}, \mathbb{P})$. That is, $|h|$ is an envelope of $\mathscr{F}_h$.

In general, there are measurability problems in the study of generalized empirical processes. However, these problems are always negligible if the considered function class is pointwise measurable (PM), see Kosorok (2008, Section 8.2, p. 142). The class $\mathscr{F}_h$ is PM according to Kosorok (2008, Lemma 8.10).

According to van der Vaart and Wellner (1996, Example 3.9.33), the class $\mathscr{H}$ of all half-spaces is a Vapnik-Červonenkis class (VC). The same is obviously true for the class $\mathscr{H}_{\mathbb{R}} = \{H \times \mathbb{R} : H \in \mathscr{H}\}$. Now, apply Kosorok (2008, Lemma 9.8) to get that the subgraphs of the associated indicator functions of $\mathscr{H}_{\mathbb{R}}$ are VC, wherein the subgraph of a real-valued function f defined on some set A is the set $\{(a, t) : t < f(a)\}$. This shows that $\mathscr{F}$ is VC. In addition, $\mathscr{F}_h$ is VC due to Kosorok (2008, Lemma 9.9 (vi)).

Overall, we have now seen that $\mathscr{F}_h$ is a PM VC class with envelope $|h|$ such that $\mathbb{E}(h^2) < \infty$. This shows that $\mathscr{F}_h$ is a $\mathbb{P}_{X,Y}$ Donsker class, see Kosorok (2008, last para., p. 165) and we can apply Kosorok (2008, Lemma 8.17) to get for every $\varepsilon > 0$

$$\mathbb{P}\left(\sup_{f,g\in\mathscr{F}_h:\rho(f,g)\le\delta} \left|\alpha_n(f) - \alpha_n(g)\right| > \varepsilon\right) \longrightarrow 0, \quad \text{for } \delta \to 0, \tag{6.16}$$

where

$$\rho(f, g) = \left(\mathbb{E}\left((f(X, Y) - g(X, Y))^2\right)\right)^{1/2}.$$

According to (6.16), the proof is complete if we can show that for an arbitrary $\delta > 0$

$$\sup_{u\in\mathbb{R}} \mathbb{E}\left(\left(\mathrm{I}_{\{\beta^\top X\le u\}} - \mathrm{I}_{\{\beta_0^\top X\le u\}}\right)^2 h^2(X, Y)\right) \le \delta, \quad \text{for } \beta \to \beta_0. \tag{6.17}$$

Note that

$$\mathbb{E}\left(\left(\mathrm{I}_{\{\beta^\top X\le u\}} - \mathrm{I}_{\{\beta_0^\top X\le u\}}\right)^2 h^2(X, Y)\right) = \mathbb{E}\left(\left|\mathrm{I}_{\{\beta^\top X\le u\}} - \mathrm{I}_{\{\beta_0^\top X\le u\}}\right| h^2(X, Y)\right).$$

Denote the integral on the right by $A(\beta, \beta_0, u)$. Then, for $K > 0$, we get by conditioning with respect to X

$$\begin{aligned} A(\beta, \beta_0, u) &\le \int \left|\mathrm{I}_{\{\beta^\top x\le u\}} - \mathrm{I}_{\{\beta_0^\top x\le u\}}\right| \sigma^2(x)\mathrm{I}_{\{\|x\|\le K\}}\mathbb{P}_X(\mathrm{d}x) \\ &\quad + \int \sigma^2(x)\mathrm{I}_{\{\|x\|>K\}}\mathbb{P}_X(\mathrm{d}x) \\ &= A_1(\beta, \beta_0, u, K) + A_2(K). \end{aligned}$$

Now choose an arbitrary $\gamma > 0$ and note that $|\beta^\top x - \beta_0^\top x| \le \|\beta - \beta_0\| \, \|x\|$, where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^p$, to get

$$
\begin{aligned}
A_1(\beta, \beta_0, u, K) &\le \int \left|\mathrm{I}_{\{\beta^\top x \le u\}} - \mathrm{I}_{\{\beta_0^\top x \le u\}}\right| \sigma^2(x) \mathrm{I}_{\{\|x\| \le K\}} \mathrm{I}_{\{|\beta^\top x - \beta_0^\top x| \le \gamma\}} \mathbb{P}_X(\mathrm{d}x) \\
&\quad + \mathrm{I}_{\{\|\beta - \beta_0\| > \gamma/K\}} \int \sigma^2(x) \mathbb{P}_X(\mathrm{d}x) \\
&\le \int \left(\mathrm{I}_{\{\beta_0^\top x \le u+\gamma\}} - \mathrm{I}_{\{\beta_0^\top x \le u-\gamma\}}\right) \sigma^2(x) \mathrm{I}_{\{\|x\| \le K\}} \mathrm{I}_{\{|\beta^\top x - \beta_0^\top x| \le \gamma\}} \mathbb{P}_X \mathrm{d}x) \\
&\quad + \mathrm{I}_{\{\|\beta - \beta_0\| > \gamma/K\}} \int \sigma^2(x) \mathbb{P}_X(\mathrm{d}x) \\
&\le \Big(H(u+\gamma, \beta_0) - H(u-\gamma, \beta_0)\Big) + \mathrm{I}_{\{\|\beta - \beta_0\| > \gamma/K\}} \int \sigma^2(x) \mathbb{P}_X(\mathrm{d}x) \\
&= A_{1,1}(\beta_0, u, \gamma) + A_{1,2}(\beta, \beta_0, \gamma, K).
\end{aligned}
$$

Overall, we have that

$$
A(\beta, \beta_0, u) \le A_{1,1}(\beta_0, u, \gamma) + A_{1,2}(\beta, \beta_0, \gamma, K) + A_2(K).
$$

Since $\mathbb{E}(\sigma^2(X)) < \infty$, we can find a $K > 0$ such that $A_2(K) < \delta$. By assumption 6.4.3(vi), H is uniformly continuous in u at β_0 and we therefore can find a $\gamma > 0$ such that for a given $\delta > 0$, $\sup_{|u-v| \le 2\gamma} |H(u, \beta_0) - H(v, \beta_0)| \le \delta$. In conclusion, if we take $\|\beta - \beta_0\| < \min(2\gamma, \gamma/K)$ we get

$$
\sup_{u \in \mathbb{R}} \mathbb{E}\left(\left(\mathrm{I}_{\{\beta^\top X \le u\}} - \mathrm{I}_{\{\beta_0^\top X \le u\}}\right)^2 h^2(X, Y)\right) \le 2\,\delta,
$$

which completes the proof of the lemma. □

Remark 6.21 The proof of the last lemma has shown that $\mathscr{H}$ is a PM VC class. Thus, it is also a Glivenko-Cantelli (GC) class, that is, w.p.1

$$
\sup_{H \in \mathscr{H}} \left| 1/n \sum_{i=1}^{n} \mathrm{I}_{\{X_i \in H\}} - \int \mathrm{I}_{\{X \in H\}} \mathrm{d}\mathbb{P} \right| \longrightarrow 0, \text{ as } n \to \infty.
$$

Interpret $\mathscr{H}$ as the class of indicator functions based on the half-spaces and multiply each indicator by a function w, such that $\mathbb{E}(|w(X)|) < \infty$, then Kosorok (2008, Corollary 9.27) guarantees that

$$
\sup_{H \in \mathscr{H}} \left| 1/n \sum_{i=1}^{n} w(X_i) \mathrm{I}_{\{X_i \in H\}} - \int w(X) \mathrm{I}_{\{X \in H\}} \mathrm{d}\mathbb{P} \right| \longrightarrow 0, \text{ as } n \to \infty,
$$

w.p.1.

The main result of this section is the following theorem, compare Stute and Zhu (2002, Theorem1).

Theorem 6.22 *Assume $\mathbb{E}(Y^2) < \infty$, F_{β_0}, the distribution function of $\beta_0^\top X$, is continuous, conditions 6.4.3(i) – 6.4.3(vi) are met, and $m(x) = m(\beta_0^\top x, \theta_0)$ holds for* *all $x \in \mathbb{R}^p$. Then, $\bar{R}_n^1$ converges in $D[-\infty, \infty]$ to a centered Gaussian process $\bar{R}_\infty^1 = R_\infty - W^\top V$, where V is a centered $(p+q)-$dimensional normal vector with covariance $L(\vartheta_0)$, W is defined in (6.15), and R_∞ is a centered Gaussian process with covariance function*

$$K(s,t) = \mathbb{E}\big(R_\infty(s)\, R_\infty(t)\big) = \int I_{\{u \le s \wedge t\}} \sigma^2_{\vartheta_0}(u)\, F_{\beta_0}(\mathrm{d}u).$$

The covariance between R_∞ and $W^\top V$ is given by

$$COV(R_\infty(s), W^\top(t)V) = W^\top(t)\,\mathbb{E}\left(\big(Y - m(\beta_0^\top X, \theta_0)\big) l(X, Y, \vartheta_0) I_{\{\beta_0^\top X \le s\}}\right), \tag{6.18}$$

and the covariance function of $\bar{R}_\infty^1$ by

$$\begin{aligned}\bar{K}^1(s,t) = K(s,t) &+ W^\top(s) L(\vartheta_0) W(t) \\ &- W^\top(s)\mathbb{E}(I_{\{\beta_0^\top X \le t\}}(Y - m(\beta_0^\top X, \theta_0)) l(X, Y, \vartheta_0)) \\ &- W^\top(t)\mathbb{E}(I_{\{\beta_0^\top X \le s\}}(Y - m(\beta_0^\top X, \theta_0)) l(X, Y, \vartheta_0)).\end{aligned} \tag{6.19}$$

Proof As the proof will show, the EMEPE is stochastically equivalent to an EMEP in which the estimated direction of evolution, β_n, is replaced by the true direction β_0. To do this, we first define the associated EMEP R_n^1

$$R_n^1(u) = n^{-1/2} \sum_{i=1}^n \big(Y_i - m(X_i, \vartheta_n)\big) \mathrm{I}_{\{\beta_0^\top X_i \le u\}}, \quad \text{for } u \in [-\infty, \infty].$$

Note that we change the notation $m(\beta^\top X, \theta)$ back to $m(X, \vartheta)$, since the subsequent proof is closely based on Theorem 6.15 and this theorem uses the notation $m(X, \vartheta)$. Related to R_n^1 is the BMEP R_n, which is defined by

$$R_n(u) = n^{-1/2} \sum_{i=1}^n \big(Y_i - m(X_i, \vartheta_0)\big) \mathrm{I}_{\{\beta_0^\top X_i \le u\}}, \quad \text{for } u \in [-\infty, \infty].$$

Due to (6.10), R_n is a BMEP with respect to the input $\beta_0^\top X_1, \ldots, \beta_0^\top X_n$. Since F_{β_0} is continuous and $\mathbb{E}(Y^2) < \infty$, we can apply Corollary 6.10 and Theorem 6.8 to get that R_n tends in distribution to the centered Gaussian process R_∞ in $D[-\infty, \infty]$ with covariance function

$$K(s,t) = \int \mathrm{I}_{\{u \le s \wedge t\}} \sigma^2_{\vartheta_0}(u)\, F_{\beta_0}(\mathrm{d}u).$$

The asymptotics of EMEP R_n^1 can be obtained as in the proof from Theorem 6.15 and we derive that R_n^1 tends in distribution in $D[-\infty, \infty]$ to a centered Gaussian process R_∞^1 with covariance function

$$K^1(s,t) = K(s,t) + W^\top(s,\vartheta_0)L(\vartheta_0)W(t,\vartheta_0)$$
$$- W^\top(s,\vartheta_0)\mathbb{E}(\mathrm{I}_{\{\beta_0^\top X\le t\}}(Y - m(X,\vartheta_0))l(X,Y,\vartheta_0))$$
$$- W^\top(t,\vartheta_0)\mathbb{E}(\mathrm{I}_{\{\beta_0^\top X\le s\}}(Y - m(X,\vartheta_0))l(X,Y,\vartheta_0)).$$

Now take the true error $e_i = Y_i - m(\beta_0^\top X_i, \theta_0)$ and split $R_n^1(u) - \bar{R}_n^1(u)$ as follows:

$$\begin{aligned} R_n^1(u) - \bar{R}_n^1(u) &= n^{-1/2}\sum_{i=1}^n e_i\big(\mathrm{I}_{\{\beta_0^\top X_i\le u\}} - \mathrm{I}_{\{\beta_n^\top X_i\le u\}}\big) \\ &\quad + n^{-1/2}\sum_{i=1}^n \Big(m(\beta_n^\top X_i,\theta_n) - m(\beta_0^\top X_i,\theta_0)\Big)\big(\mathrm{I}_{\{\beta_n^\top X_i\le u\}} - \mathrm{I}_{\{\beta_0^\top X_i\le u\}}\big) \\ &= A_1(\beta_n,\vartheta_0,u) + A_2(\vartheta_n,\vartheta_0,u). \end{aligned}$$

According to Lemma 6.20, since $\vartheta_n \to \vartheta_0$ in probability,

$$\sup_{u\in\mathbb{R}} |A_1(\beta_n,\vartheta_0,u)| \longrightarrow 0, \quad \text{as } n\to\infty,$$

in probability. For $A_2(\vartheta_n,\vartheta_0,u)$, we use a Taylor expansion and derive by Lemma 6.14, similar as in the proof of Theorem 6.15, that

$$A_2(\vartheta_n,\vartheta_0,u) = n^{1/2}\big(\vartheta_n - \vartheta_0\big)^\top n^{-1}\sum_{i=1}^n w(X_i,\vartheta_0)\big(\mathrm{I}_{\{\beta_n^\top X_i\le u\}} - \mathrm{I}_{\{\beta_0^\top X_i\le u\}}\big) + o_{\mathbb{P}}(1),$$

uniformly in u, as $n\to\infty$. Since $n^{1/2}\big(\vartheta_n - \vartheta_0\big)$ tends to a normal distribution, it remains to show that

$$\sup_{u\in\mathbb{R}}\Big|n^{-1}\sum_{i=1}^n w_k(X_i,\vartheta_0)\big(\mathrm{I}_{\{\beta_n^\top X_i\le u\}} - \mathrm{I}_{\{\beta_0^\top X_i\le u\}}\big)\Big| \longrightarrow 0, \quad \text{as } n\to\infty,$$

in probability, for $1\le k\le p+q$. For this, assume that $|\beta_n - \beta_0| < \gamma$ for $\gamma > 0$. Then, for $1\le k\le p+q$ the supremum is bounded from above by

$$\begin{aligned} &\sup_{u\in\mathbb{R},|\beta-\beta_0|<\gamma}\Big|n^{-1}\sum_{i=1}^n w_k(X_i,\vartheta_0)\mathrm{I}_{\{\beta^\top X_i\le u\}} - \mathbb{E}\big(w_k(X,\vartheta_0)\mathrm{I}_{\{\beta^\top X\le u\}}\big)\Big| \\ &\quad + \Big|n^{-1}\sum_{i=1}^n w_k(X_i,\vartheta_0)\mathrm{I}_{\{\beta_0^\top X_i\le u\}} - \mathbb{E}\big(w_k(X,\vartheta_0)\mathrm{I}_{\{\beta_0^\top X\le u\}}\big)\Big| \\ &\quad + \sup_{u\in\mathbb{R},|\beta-\beta_0|<\gamma}\mathbb{E}\big(M(X)|\mathrm{I}_{\{\beta^\top X\le u\}} - \mathrm{I}_{\{\beta_0^\top X\le u\}}|\big). \end{aligned}$$

According to Remark 6.21, the first two terms in the above bound tend to 0. Since $\beta_n \to \beta_0$ in probability, as $n\to\infty$, the third term tends to 0 with the same argumentation that was used to prove (6.17). This finally completes the proof of the theorem. □

6.5 Mathematical Framework: Bootstrap of Marked Empirical Processes

This section examines the asymptotics of bootstrap variants of the BMEP and EMEP processes in the GLM context. At first, this is done, in general, without using a special resampling method and without distinguishing between parametric and semi-parametric models. Specifications of the theoretical results obtained with respect to these two concrete models are given at the end of this section.

Let us recall Definition 6.17 of the EMEPE $\bar{R}_n^1$ and the proof of Theorem 6.22. There it was shown that $\bar{R}_n^1$ is asymptotically equivalent to the particular EMEP

$$R_n^1(u) = n^{-1/2} \sum_{i=1}^{n} \big(Y_i - m(\beta_n^\top X_i, \theta_n)\big) \mathrm{I}_{\{\beta_0^\top X_i \le u\}}.$$

The corresponding bootstrap analog to this EMEP is given in the following definition.

Definition 6.23 Assume the setup of Definition 6.17, let $(Y_{1,n}^*, X_1), \ldots, (Y_{n,n}^*, X_n)$ be the bootstrap data according to some resampling scheme such that $\mathbb{E}_n^*(Y_{i,n}^*) = m(X_i, \vartheta_n)$, where $\vartheta_n = (\beta_n, \theta_n)$ is the estimate of ϑ_0 based on the original data. Denote with $\vartheta_n^* = (\beta_n^*, \theta_n^*)$ the estimated parameter based on the bootstrap data. Then

$$[-\infty, \infty] \ni u \longrightarrow R_n^{1*}(u) = n^{-1/2} \sum_{i=1}^{n} \big(Y_{i,n}^* - m(\beta_n^{*\top} X_i, \theta_n^*)\big) \mathrm{I}_{\{\beta_n^\top X_i \le u\}} \tag{6.20}$$

defines the *bootstrapped estimated marked empirical process*.

Remark 6.24 For a direct transfer of EMEPE into the bootstrap world, instead of the indicator $\mathrm{I}_{\{\beta_n^\top X_i \le u\}}$, one would have to actually use the indicator $\mathrm{I}_{\{\beta_n^{*\top} X_i \le u\}}$ in Definition 6.23. But, as already noted, EMEPE and EMEP are stochastically equivalent for the original data. Furthermore, the bootstrap version of EMEP has the big advantage that in Monte Carlo simulations to determine corresponding statistics, the values of $\beta_n^\top X_1, \ldots, \beta_n^\top X_n$ only have to be sorted once and not separately for each individual bootstrap dataset, which would be necessary in the case of the indicator $\mathrm{I}_{\{\beta_n^{*\top} X_i \le u\}}$. Due to this performance advantage, we have only considered the EMEP variant for the bootstrap procedure here.

Note that

$$\begin{aligned} R_n^{1*}(u) &= n^{-1/2} \sum_{i=1}^{n} \left(Y_{i,n}^* - m(\beta_n^{*\top} X_i, \theta_n^*)\right) \mathrm{I}_{\{\beta_n^\top X_i \le u\}} \\ &= n^{-1/2} \sum_{i=1}^{n} \left(Y_{i,n}^* - m(\beta_n{}^\top X_i, \theta_n)\right) \mathrm{I}_{\{\beta_n^\top X_i \le u\}} \\ &\quad - n^{-1/2} \sum_{i=1}^{n} \left(m(\beta_n^{*\top} X_i, \theta_n^*) - m(\beta_n{}^\top X_i, \theta_n)\right) \mathrm{I}_{\{\beta_n^\top X_i \le u\}} \\ &= R_n^*(u) - S_n^*(u). \end{aligned}$$

Within the bootstrap world, the first process on the right side, that is,

$$R_n^*(u) = n^{-1/2} \sum_{i=1}^{n} \left(Y_{i,n}^* - m(\beta_n{}^\top X_i, \theta_n)\right) \mathrm{I}_{\{\beta_n^\top X_i \le u\}}, \tag{6.21}$$

can be interpreted as a bootstrap version of a BMEP, since $\mathbb{E}_n^*(Y_{i,n}^*) = m(\beta_n{}^\top X_i, \theta_n)$.

The second process

$$S_n^*(u) = n^{-1/2} \sum_{i=1}^{n} \left(m(\beta_n^{*\top} X_i, \theta_n^*) - m(\beta_n{}^\top X_i, \theta_n)\right) \mathrm{I}_{\{\beta_n^\top X_i \le u\}} \tag{6.22}$$

deals with the influence of parameter estimation in m.

The two bootstrap methods considered so far have two things in common. First, the X_i from the underlying dataset is taken directly into the bootstrap dataset, i.e., $X_{i,n}^* = X_i$. So $X_{i,n}^*$ in the bootstrap dataset is deterministic and not random like in the original dataset! Second, the corresponding $Y_{i,n}^*$ has the property $\mathbb{E}_n^*(Y_{i,n}^*) = m(\beta_n^\top X_i, \theta_n)$.

The goal in this chapter is to prove that w.p.1, R_n^{1*} converges toward the same limit process $\bar{R}_\infty^1$ as $\bar{R}_n^1$ does.

To get a more compact notation, we will write $m(x, \vartheta)$ for $m(\beta^\top x, \theta)$ in different places, where $\vartheta = (\beta, \theta)$.

In the forthcoming proofs, we will base ourselves on arguments which require a special condition and which we now summarize in advance in the following definition.

Definition 6.25 Let V be a compact neighborhood of ϑ_0 and

$$h: \quad \mathbb{R}^{p+1} \times V \ni (x, y, \vartheta) \longrightarrow h(x, y, \vartheta) \in \mathbb{R}$$

a measurable function such that $h(x, y, \vartheta)$ is continuous in $\vartheta = (\beta, \theta)$ for all $\vartheta \in V$ and $(x, y) \in \mathbb{R}^{p+1}$. We call such a function h *uniformly dominated by M over V at*

ϑ_0 if there exists a $\mathbb{P}_{X,Y}-$ integrable function M such that $|h(x, y, \vartheta)| \leq M(x, y)$ for all $\vartheta \in V$ and $(x, y) \in \mathbb{R}^{p+1}$.

In the following two sections, we will sometimes use a technical argument in the proofs, which we next formulate here as a lemma. The proof of this lemma is based in its technique on the proof of Lemma 6.20.

Lemma 6.26 *Let h be uniformly dominated by M over V at ϑ_0 and assume that*

$$H : \mathbb{R} \ni u \longrightarrow H(u) = \mathbb{E}\big(|h(X, Y, \vartheta_0)| I_{\{\beta_0^\top X \leq u\}}\big) \in \mathbb{R}$$

is uniformly continuous in u. Then we get

(i) As $n \to \infty$,

$$\sup_{\vartheta \in V,\, u \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{n} h(X_i, Y_i, \vartheta) I_{\{\beta^\top X_i \leq u\}} - \mathbb{E}\big(h(X, Y, \vartheta) I_{\{\beta^\top X \leq u\}}\big) \right| \longrightarrow 0, \quad \textit{w.p.1.}$$

(ii) As $\varepsilon \to 0$,

$$\sup_{\|\vartheta - \vartheta_0\| \leq \varepsilon,\, u \in \mathbb{R}} \left| \mathbb{E}\big(h(X, Y, \vartheta) I_{\{\beta^\top X \leq u\}}\big) - \mathbb{E}\big(h(X, Y, \vartheta_0) I_{\{\beta_0^\top X \leq u\}}\big) \right| \longrightarrow 0.$$

(iii) If $\vartheta_n \to \vartheta_0$ w.p.1, then, as $n \to \infty$,

$$\sup_{u \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{n} h(X_i, Y_i, \vartheta_n) I_{\{\beta_n^\top X_i \leq u\}} - \mathbb{E}\big(h(X, Y, \vartheta_0) I_{\{\beta_0^\top X \leq u\}}\big) \right| \longrightarrow 0,$$

w.p.1.

Proof Theorem 5.66 guarantees that

$$\mathscr{G} = \big\{ h(\cdot, \cdot, \vartheta) \,:\, \vartheta \in V \big\}$$

is a PM-GC class (pointwise measurable Glivenko-Cantelli class) with integrable envelope M. As already pointed out in the proof of Lemma 6.20, the collection

$$\mathscr{F} = \big\{ \mathrm{I}_{\{H \times \mathbb{R}\}} \,:\, H \text{ is half-space in } \mathbb{R}^p \big\}$$

is a PM-VC class and therefore a PM-GC class. In summary, we then get from Kosorok (2008, Corollary 9.27) that

$$\mathscr{F} = \big\{ h(\cdot, \cdot, \vartheta) \mathrm{I}_{\{\beta^\top \cdot \leq u\}} \,:\, (\beta, \theta) = \vartheta \in V \text{ and } u \in \mathbb{R} \big\}$$

is a PM-GC class which completes the proof of part (i) of the lemma.

For (ii), first observe that

$$\begin{aligned}
\sup_{\|\vartheta-\vartheta_0\|\le\varepsilon,\, u\in\mathbb{R}} &\Big|\mathbb{E}\big(h(X,Y,\vartheta)\mathrm{I}_{\{\beta^\top X\le u\}}\big) - \mathbb{E}\big(h(X,Y,\vartheta_0)\mathrm{I}_{\{\beta_0^\top X\le u\}}\big)\Big| \\
&\le \mathbb{E}\Big(\sup_{\|\vartheta-\vartheta_0\|\le\varepsilon} \big|h(X,Y,\vartheta)-h(X,Y,\vartheta_0)\big|\Big) \\
&\quad + \sup_{\|\beta-\beta_0\|<\varepsilon,\, u\in\mathbb{R}} \mathbb{E}\Big(\big|h(X,Y,\vartheta_0)\big|\,\big|\mathrm{I}_{\{\beta^\top X\le u\}} - \mathrm{I}_{\{\beta_0^\top X\le u\}}\big|\Big).
\end{aligned}$$

The assumptions, according to the dominated convergence theorem, guarantee that the first term on the right side converges to 0 as $\varepsilon \to 0$. Denote the expectation appearing in the second term on the right side by $A(\beta, u, \varepsilon)$ and choose $K > 0$ to get

$$\begin{aligned}
A(\beta,u,\varepsilon) &\le \mathbb{E}\Big(\big|h(X,Y,\vartheta_0)\big|\,\big|\mathrm{I}_{\{\beta^\top X\le u\}} - \mathrm{I}_{\{\beta_0^\top X\le u\}}\big|\mathrm{I}_{\{\|X\|\le K\}}\Big) \\
&\quad + \mathbb{E}\Big(\big|h(X,Y,\vartheta_0)\big|\mathrm{I}_{\{\|X\|>K\}}\Big) \\
&= A_1(\beta,u,\varepsilon,K) + A_2(K).
\end{aligned}$$

Next select $\gamma > 0$ to get

$$\begin{aligned}
A_1(\beta,u,\varepsilon,K) &\le \mathbb{E}\Big(\big|h(X,Y,\vartheta_0)\big|\,\big|\mathrm{I}_{\{\beta^\top X\le u\}} - \mathrm{I}_{\{\beta_0^\top X\le u\}}\big|\mathrm{I}_{\{\|X\|\le K\}}\mathrm{I}_{\{|\beta^\top X-\beta_0^\top X|\le\gamma\}}\Big) \\
&\quad + \mathrm{I}_{\{\|\beta-\beta_0\|>\gamma/K\}}\mathbb{E}\Big(\big|h(X,Y,\vartheta_0)\big|\Big) \\
&\le \mathbb{E}\Big(\big|h(X,Y,\vartheta_0)\big|\,\big|\mathrm{I}_{\{\beta_0^\top X\le u+\gamma\}} - \mathrm{I}_{\{\beta_0^\top X\le u-\gamma\}}\big|\Big) \\
&\quad + \mathrm{I}_{\{\|\beta-\beta_0\|>\gamma/K\}}\mathbb{E}\Big(\big|h(X,Y,\vartheta_0)\big|\Big) \\
&= \big(H(u+\gamma)-H(u-\gamma)\big) + \mathrm{I}_{\{\|\beta-\beta_0\|>\gamma/K\}}\mathbb{E}\Big(\big|h(X,Y,\vartheta_0)\big|\Big) \\
&= A_{1,1}(u,\gamma) + A_{1,2}(\beta,\gamma,K).
\end{aligned}$$

All in all, this means that we have

$$A(\beta,u,\varepsilon) \le A_{1,1}(u,\gamma) + A_{1,2}(\beta,\gamma,K) + A_2(K).$$

Now fix $\delta > 0$. Since $\mathbb{E}(|h(X,Y,\vartheta_0)|) < \infty$, we can find a $K > 0$ such that $A_2(K) < \delta$. H is uniformly continuous and we can therefore find a $\gamma > 0$ such that, uniformly in u, $A_{1,1} < \delta$. If we take $\varepsilon < \min(\gamma, \gamma/K)$, then $A_{1,2}(\beta,\gamma,K) = 0$, and we get for such an ε that $A(\beta,u,\varepsilon) < 2\,\delta$. This shows that

$$\sup_{\|\beta-\beta_0\|<\varepsilon,\, u\in\mathbb{R}} A(\beta,u,\varepsilon) \longrightarrow 0,$$

as $\beta \to \beta_0$. Overall, this proves part (ii) of the lemma.

Since $\vartheta_n \longrightarrow \vartheta_0$ w.p.1, part (iii) now follows directly from (i) and (ii). This completes the proof of the lemma. □

6.5.1 Bootstrap of the BMEP

To prove a corresponding functional limit theorem for the bootstrap version of the BMEP we will make use of the following assumptions.

(i) $\mathbb{E}_n^*(Y_{i,n}^*) = m(\beta_n^\top X_i, \theta_n) \equiv m(X_i, \vartheta_n)$, where we set as before $\vartheta_n = (\beta_n, \theta_n)$. There exists a $\delta > 0$ and a non-negative function

$$h_e : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^p \times \Theta \ni (x, y, \beta, \theta) \equiv (x, y, \vartheta) \longrightarrow h_e(x, y, \vartheta) \in \mathbb{R}$$

such that $\mathbb{E}_n^*(|Y_{i,n}^* - m(X_i, \vartheta_n)|^{2+\delta}) = h_e(X_i, Y_i, \vartheta_n)$ and h_e is uniformly dominated by M_e over V at ϑ_0 for some function M_e and a compact neighborhood V of ϑ_0.

(ii) There exists a non-negative function

$$h_v : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^p \times \Theta \ni (x, y, \beta, \theta) \equiv (x, y, \vartheta) \longrightarrow h_v(x, y, \vartheta) \in \mathbb{R}$$

such that $\mathrm{VAR}_n^*(Y_{i,n}^*) = h_v(X_i, Y_i, \vartheta_n)$ and for all $u \in \mathbb{R}$

$$\mathbb{E}\big(h_v(X, Y, \vartheta_0)\mathrm{I}_{\{\beta_0^\top X \le u\}}\big) = \int \mathrm{I}_{\{t \le u\}} \sigma_{\vartheta_0}^2(t)\, F_{\beta_0}(\mathrm{d}t),$$

where F_{β_0} is the distribution function of $\beta_0^\top X$ and $\sigma_{\vartheta_0}^2$ is defined under (6.14). Furthermore, h_v is uniformly dominated by M_v over V at ϑ_0 for some function M_v and a compact neighborhood V of ϑ_0 and the function

$$H_v : \mathbb{R} \times V \ni (u, \beta, \theta) \equiv (u, \vartheta) \longrightarrow H_v(u, \vartheta) := \mathbb{E}\big(\mathrm{I}_{\{\beta^\top X \le u\}} h_v(X, Y, \vartheta)\big)$$

is uniformly continuous in u at ϑ_0.

(iii) $\vartheta_n \longrightarrow \vartheta_0$, as $n \to \infty$, w.p.1.

The first two conditions seem somewhat unusual at first glance. They specify conditions for the bootstrap moments, which depend on the respective resampling procedure. Thus, we can treat the two resampling methods with only one theorem.

Theorem 6.27 *Assume that conditions 6.5.1(i), 6.5.1(ii), and 6.5.1(iii) are met. Then, w.p.1, the process R_n^* converges in $D[-\infty, \infty]$ to a centered Gaussian process R_∞ with covariance function*

$$K(s, t) = \mathbb{E}\big(R_\infty(s)\, R_\infty(t)\big) = \int I_{\{u \le s \wedge t\}} \sigma_{\vartheta_0}^2(u)\, F_{\beta_0}(du).$$

Proof To prove the assertion we use again Billingsley (1968, Theorem 15.6).

For this, we first show that the fidis of R_n^* converge in distribution to those of R_∞. Let $k \in \mathbb{N}$ and take $-\infty \le t_1 < \ldots < t_k \le \infty$. According to Cramér-Wold, see Billingsley (1968, Theorem 7.7), we have to show that, w.p.1, for every $0 \ne a \in \mathbb{R}^k$

$$\sum_{j=1}^{k} a_j R_n^*(t_j) \longrightarrow \mathcal{N}(0, a^\top \Sigma a), \quad \text{as } n \to \infty,$$

in distribution, where $\Sigma = (\sigma_{j,\ell})_{1\le j,\ell\le k}$ and $\sigma_{j,\ell} = \mathrm{COV}(R_\infty(t_j), R_\infty(t_\ell)) = K(t_j, t_\ell)$.

Set

$$\begin{aligned} Z_n^* = \sum_{j=1}^{k} a_j R_n^*(t_j) &= n^{-1/2} \sum_{i=1}^{n} \Big(\big(Y_{i,n}^* - m(X_i, \vartheta_n)\big)\big(\sum_{j=1}^{k} a_j \mathrm{I}_{\{\beta_n^\top X_i \le t_j\}}\big)\Big) \\ &\equiv \sum_{i=1}^{n} \xi_{i,n}^* A_{i,n}, \end{aligned}$$

where $\xi_{i,n}^* = n^{-1/2}\big(Y_{i,n}^* - m(X_i, \vartheta_n)\big)$. Note that $\xi_{1,n}^*, \ldots, \xi_{n,n}^*$ are independent and centered, because of 6.5.1(i). Furthermore, $A_{1,n}, \ldots, A_{n,n}$ are deterministic with respect to $\mathbb{P}_n^*$.

We first consider the variance of Z_n^* and get

$$\begin{aligned} \mathrm{VAR}_n^*(Z_n^*) &= \sum_{1\le j,\ell\le k} a_j \Big(1/n \sum_{i=1}^{n} \mathrm{I}_{\{\beta_n^\top X_i \le t_j \wedge t_\ell\}} \mathrm{VAR}_n^*(Y_{i,n}^*)\Big) a_\ell \\ &= \sum_{1\le j,\ell\le k} a_j \Big(\frac{1}{n} \sum_{i=1}^{n} \mathrm{I}_{\{\beta_n^\top X_i \le t_j \wedge t_\ell\}} h_v(X_i, Y_i, \vartheta_n)\Big) a_\ell, \end{aligned}$$

where the last equality follows from 6.5.1(ii). Now apply Lemma 6.26 to get that w.p.1

$$\frac{1}{n} \sum_{i=1}^{n} \mathrm{I}_{\{\beta_n^\top X_i \le t_j \wedge t_\ell\}} h_v(X_i, Y_i, \vartheta_n) \longrightarrow \int \mathrm{I}_{\{u \le t_j \wedge t_\ell\}} \sigma_{\vartheta_0}^2(u)\, F_{\beta_0}(\mathrm{d}u) = K(t_j, t_\ell),$$

as $n \to \infty$, that is,

$$\mathrm{VAR}_n^*(Z_n^*) \longrightarrow a^\top \Sigma a, \quad \text{as } n \to \infty, \text{ w.p.1.} \tag{6.23}$$

Since Σ is positive semi-definite, $a^\top \Sigma a \ge 0$. If $a^\top \Sigma a = 0$, Chebyshev's inequality guarantees that $Z_n^* = o_{\mathbb{P}_n^*}(1)$ and we have that w.p.1,

$$Z_n^* \longrightarrow \mathcal{N}(0, a^\top \Sigma a), \quad \text{as } n \to \infty.$$

In this case, $\mathcal{N}(0, a^\top \Sigma a) = \mathcal{N}(0,0)$ is a degenerated normal distribution. Now assume that $a^\top \Sigma a > 0$. To prove the asymptotic normality of Z_n^*, Serfling (1980, Corollary to Theorem 1.9.3) can be applied and we have to show that the Lyapunov condition

$$\frac{1}{\mathrm{VAR}_n^*(Z_n^*)^{(2+\nu)/2}} \sum_{i=1}^n \mathbb{E}_n^*\Big(\big|\xi_{i,n}^* A_{i,n}\big|^{2+\nu}\Big) \longrightarrow 0, \quad \text{as } n \to \infty \tag{6.24}$$

is fulfilled w.p.1, for some $\nu > 1$, where the null set does not depend on a.

Since $a^\top \Sigma a > 0$, (6.23), and $|A_{i,n}| \le \|a\|\, k$, the Lyapunov condition (6.24) is, therefore, fulfilled if we can prove that

$$\begin{aligned}\sum_{i=1}^n \mathbb{E}_n^*\Big(\big|\xi_{i,n}^*\big|^{2+\delta}\Big) &= \frac{1}{n^{1+\delta/2}} \sum_{i=1}^n \mathbb{E}_n^*\Big(\big(Y_{i,n}^* - m(X_i, \vartheta_n)\big)^{2+\delta}\Big) \\ &= \frac{1}{n^{\delta/2}} \frac{1}{n} \sum_{i=1}^n h_e(X_i, Y_i, \vartheta_n) \\ &\longrightarrow 0, \quad \text{as } n \to \infty, \text{ w.p.1},\end{aligned}$$

where $\delta > 0$ and h_e are chosen according to assumption 6.5.1(i). The assumed properties of h_e together with the w.p.1 convergence $\vartheta_n \to \vartheta_0$ now yield according to Theorem 5.66

$$\sup_{\vartheta \in V} \Big| \frac{1}{n} \sum_{i=1}^n h_e(X_i, Y_i, \vartheta) - \mathbb{E}(h_e(X, Y, \vartheta_0)) \Big| \longrightarrow 0, \quad \text{as } n \to \infty, \text{ w.p.1},$$

where V is chosen according to 6.5.1(i). This proves the convergence of the finite-dimensional distributions against $\mathcal{N}(0, \Sigma)$.

It remains to show that $(R_n^*)_{n\in\mathbb{N}}$ is tight. Since $D[-\infty, \infty]$ can be identified with $D[0,1]$, compare Remark 6.4, we can adjust Billingsley (1968, Theorem 15.6) accordingly to prove tightness. For this let $-\infty \le u_1 \le u \le u_2 \le \infty$. As in the proof of Theorem 6.8, we get from Lemma 6.26 that w.p.1

$$\begin{aligned}&\limsup_{n\to\infty} \mathbb{E}_n^*\Big(\big(R_n^*(u) - R_n^*(u_1)\big)^2 \big(R_n^*(u_2) - R_n^*(u)\big)^2\Big) \\ &\quad\le \limsup_{n\to\infty} \Big(\frac{1}{n} \sum_{i=1}^n \mathbb{E}_n^*\big((Y_{i,n}^* - m(X_i, \vartheta_n))^2 \mathrm{I}_{\{u_1 < \beta_n^\top X_i \le u_2\}}\big)\Big)^2 \\ &\quad= \limsup_{n\to\infty} \Big(\frac{1}{n} \sum_{i=1}^n \mathbb{E}_n^*\big(h_v(X_i, Y_i, \vartheta_n) \mathrm{I}_{\{u_1 < \beta_n^\top X_i \le u_2\}}\big)\Big)^2 \\ &\quad= \big(H_v(u_2, \vartheta_0) - H_v(u_1, \vartheta_0)\big)^2.\end{aligned}$$

Note that $H_v(\cdot, \vartheta_0)$ is uniformly continuous.

With a small adaptation in the proof of Billingsley (1968, Theorem 15.6), more precisely under Billingsley (1968, (15.30) in the proof of Theorem 15.6), this last result now yields that, with w.p.1, $(R_n^*)_{n\in\mathbb{N}}$ is tight. □

6.5.2 Bootstrap of the EMEP

Additionally to the assumptions 6.5.1(i), 6.5.1(ii), and 6.5.1(iii) of Section 6.5.1, we need further conditions to handle the process S_n^*.

(iv) $n^{1/2}(\vartheta_n^* - \vartheta_n) \longrightarrow Z$ in distribution, as $n \to \infty$, w.p.1, where Z is a zero mean multivariate distribution with covariance matrix $L(\vartheta_0)$.

(v) $L(\vartheta_0) = \mathbb{E}(l(X, Y, \vartheta_0)l^\top(X, Y, \vartheta_0))$ exists and is positive definite.

(vi) $n^{1/2}(\vartheta_n^* - \vartheta_n) = n^{-1/2}\sum_{i=1}^n l(X_i, Y_{i,n}^*, \vartheta_n) + o_{\mathbb{P}_n^*}(1)$, as $n \to \infty$, w.p.1.

(vii) $\mathbb{E}_n^*(l(X_i, Y_{i,n}^*, \vartheta_n)) = 0$ and there exists a $\delta > 0$ and a non-negative function

$$h_{l,e} : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^p \times \Theta \ni (x, y, \beta, \theta) \equiv (x, y, \vartheta) \longrightarrow h_{l,e}(x, y, \vartheta) \in \mathbb{R}$$

such that $\mathbb{E}_n^*(\|l(X_i, Y_{i,n}^*, \vartheta_n)\|^{2+\delta}) = h_{l,e}(X_i, Y_i, \vartheta_n)$. Furthermore, $h_{l,e}$ is uniformly dominated over V at ϑ_0 for some function $M_{l,e}$ and a compact neighborhood V at ϑ_0.

(viii) The covariance matrix

$$L_n^*(\vartheta_n) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_n^*\big(l(X_i, Y_{i,n}^*, \vartheta_n)l^\top(X_i, Y_{i,n}^*, \vartheta_n)\big) \longrightarrow L(\vartheta_0), \quad \text{as } n \to \infty, \text{w.p.1.}$$

(ix) For every $x \in \mathbb{R}^p$, $w(x, \vartheta) = \partial m(x, \vartheta)/\partial\vartheta = (w_1(x, \vartheta), \ldots, w_{p+q}(x, \vartheta))^\top$ exists and is continuous with respect to ϑ for every ϑ in a neighborhood of ϑ_0 (not depending on x).

(x) For $1 \le i \le p + q$, $w_i(x, \vartheta)$ is uniformly dominated by some M_w over V at ϑ_0.

(xi) The function

$$W : \quad \mathbb{R} \times V_\beta \ni (u, \beta) \longrightarrow W(u, \beta) = \mathbb{E}\big(w(X, \vartheta_0)\mathrm{I}_{\{\beta^\top X \le u\}}\big) \in \mathbb{R}^{p+q}$$

is uniformly continuous in u at β_0, where $V_\beta = \{\beta : (\beta, \theta_0) \in V\}$ and V is given under 6.5.2(ix).

(xii) There exists a function

$$h_{cov} : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^p \times \Theta \ni (x, y, \beta, \theta) \equiv (x, y, \vartheta) \longrightarrow h_{cov}(x, y, \vartheta) \in \mathbb{R}^{p+q}$$

such that $\mathbb{E}_n^*\big((Y_{i,n}^* - m(X_i, \vartheta_n))\, l(X_i, Y_{i,n}^*, \vartheta_n)\big) = h_{cov}(X_i, Y_i, \vartheta_n)$ and for all $u \in \mathbb{R}$

$$\mathbb{E}\big(h_{cov}(X, Y, \vartheta_0)\mathrm{I}_{\{\beta_0^\top X \le u\}}\big) = \mathbb{E}\big((Y - m(\beta_0^\top X, \theta_0))\, l(X, Y, \vartheta_0)\, \mathrm{I}_{\{\beta_0^\top X \le u\}}\big).$$

Furthermore, each component $h_{cov,r}$, where $1 \le r \le p+q$, is uniformly dominated by $M_{cov,r}$ over V at ϑ_0 for some function $M_{cov,r}$ and a compact neighborhood V of ϑ_0, and the function

$$H_{cov} : \mathbb{R} \times V \ni (u, \beta, \theta) \equiv (u, \vartheta) \longrightarrow H_{cov}(u, \vartheta) = \mathbb{E}\big(\mathrm{I}_{\{\beta^\top X \le u\}} h_{cov}(X, Y, \vartheta)\big) \in \mathbb{R}^{p+q}$$

is uniformly continuous in u at ϑ_0.

In the following two remarks, we examine the validity of the moment conditions stated above for the resampling procedures used in the parametric and semi-parametric bootstraps, respectively.

Remark 6.28 According to the Resampling Scheme 5.42 for the parametric case, $Y^*_{i,n}$ has density

$$f(y|\theta_{X_i}(\beta_n), \phi_n) = \exp\left(\frac{\theta_{X_i}(\beta_n)y - \zeta(\theta_{X_i}(\beta_n))}{\phi_n}\right) h(y, \phi_n),$$

where β_n and ϕ_n are the MLE corresponding to the original dataset and $\theta_x(\beta) = (g \circ \zeta')^{-1}(\beta^\top x)$. Now apply (5.19) to get

$$\mathrm{VAR}^*_n(Y^*_{i,n}) = h_v(X_i, Y_i, \vartheta_n) = \phi_n \zeta''(\theta_{X_i}(\beta_n)).$$

This representation of h_v can be used to find conditions that guarantee the validity of assumption 6.5.1(ii). The situation is similar with assumption 6.5.1(i).

Various assumptions state that for some function u

$$\mathbb{E}^*_n\big(u(X_i, Y^*_{i,n}, \vartheta_n)\big) = h(X_i, Y_i, \vartheta_n)$$

is uniformly dominated by a function M over a compact neighborhood V of ϑ_0. In case of the parametric bootstrap, we have

$$\mathbb{E}^*_n\big(u(X_i, Y^*_{i,n}, \vartheta_n)\big) = \int u(X_i, y, \vartheta_n) f(y|\theta_{X_i}(\beta_n), \phi_n)\nu(\mathrm{d}y),$$

which is only a function of X_i and ϑ_n. Looking at the density of $Y^*_{i,n}$ reveals that one could also write it as

$$\exp\Big(K_1(X_i, \vartheta_n)y + c(y, \phi_n)\Big)\, K_2(X_i, \vartheta_n)$$

which even simplifies further, for instance, for the normal, Poisson, Bernoulli, gamma, and inverse Gaussian distribution to

$$\exp\Big(K_1(X_i, \vartheta_n)y + K_3(\phi_n)c(y)\Big)\, K_2(X_i, \vartheta_n).$$

Therefore, if $\|X_i\|$ is bounded by some K, one can try to shrink the neighborhood V and obtain the bound

$$\mathbb{E}_n^*\big(u(X_i, Y_{i,n}^*, \vartheta_n)\big) \le \int \mathrm{I}_{\{\|X_i\|<K\}} u(X_i, y, \vartheta_n) \exp\Big(C_1 y + C_3 c(y)\Big)\, C_2 \nu(\mathrm{d}y).$$

On the basis of such a representation, conditions can then be formulated that ultimately guarantee the required integrability conditions.

Finally, note that Corollary 5.62 already provides a linear expansion and assumption 6.5.2(vi) and 6.5.2(viii) as well as $\mathbb{E}_n^*(l(X_i, Y_{i,n}^*, \vartheta_n)) = 0$ from assumption 6.5.2(vii).

Remark 6.29 According to the Resampling scheme 5.64 a wild bootstrap is used and

$$Y_{i,n}^* = m(X_i, \vartheta_n) + \tau_i \left(Y_i - m(X_i, \vartheta_n)\right),$$

where τ_i is a Rademacher variable, that is, $\mathbb{P}(\tau = 1) = 1/2 = \mathbb{P}(\tau = -1)$, also independent of X_i and Y_i.

Various assumptions state that for some function u

$$\mathbb{E}_n^*\big(u(X_i, Y_{i,n}^*, \vartheta_n)\big) = h(X_i, Y_i, \vartheta_n)$$

is uniformly dominated by a function M over a compact neighborhood V of ϑ_0. Obviously,

$$\mathbb{E}_n^*\big(u(X_i, Y_{i,n}^*, \vartheta_n)\big) = u(X_i, Y_i, \vartheta_n)\mathrm{I}_{\{\tau_i=-1\}} + u(X_i, 2m(X_i, \vartheta_n) - Y_i, \vartheta_n)\mathrm{I}_{\{\tau_i=1\}}$$

.

On the basis of such a representation, conditions can then be formulated that ultimately guarantee the required integrability conditions.

Finally, note that Corollary 5.81 already provides a linear expansion and assumption 6.5.2(vi) and 6.5.2(viii) as well as $\mathbb{E}_n^*(l(X_i, Y_{i,n}^*, \vartheta_n)) = 0$ from assumption 6.5.2(vii).

The following lemma shows that condition 6.5.2(iv) is a consequence of conditions 6.5.1(iii), 6.5.2(v), 6.5.2(vi), 6.5.2(vii), and 6.5.2(viii). In order to obtain a more compact notation, we have retained it in the list of conditions.

Lemma 6.30 *Assume that conditions 6.5.1(iii), 6.5.2(v), 6.5.2(vi), 6.5.2(vii), and 6.5.2(viii) hold. Then, w.p.1,*

$$n^{-1/2}(\vartheta_n^* - \vartheta_n) \longrightarrow Z, \quad as\ n \to \infty,$$

where Z is multivariate normally distributed with zero mean and covariance matrix $L(\vartheta_0)$.

Proof Due to 6.5.2(vi) and Cramér-Wold, see Billingsley (1968, Theorem 7.7), we have to show that, w.p.1, for every $0 \neq a \in \mathbb{R}^{p+q}$

$$Z_n^* = n^{-1/2} \sum_{i=1}^{n} a^\top l(X_i, Y_{i,n}^*, \vartheta_n) \longrightarrow \mathcal{N}(0, a^\top L(\vartheta_0)a), \quad \text{as } n \to \infty,$$

in distribution. According to Serfling (1980, Corollary to Theorem 1.9.3), this follows if we can show that for some $\nu > 0$ the Lyapunov condition

$$\frac{1}{\mathrm{VAR}_n^*(Z_n^*)^{(2+\nu)/2}} \sum_{i=1}^{n} \mathbb{E}_n^*\Big(\big|n^{-1/2} a^\top l(X_i, Y_{i,n}^*, \vartheta_n)\big|^{2+\nu}\Big) \longrightarrow 0, \quad \text{as } n \to \infty, \text{ w.p.1}$$

holds, where the null set does not depend on a.

For the variance, we get from 6.5.2(viii)

$$\begin{aligned}\mathrm{VAR}_n^*(Z_n^*) &= \frac{1}{n} \sum_{i=1}^{n} a^\top \mathbb{E}_n^*\Big(l(X_i, Y_{i,n}^*, \vartheta_n)\, l^\top(X_i, Y_{i,n}^*, \vartheta_n)\Big)a = a^\top L_n^*(\vartheta_n)a \\ &\longrightarrow a^\top L(\vartheta_0)a,\end{aligned}$$

as $n \to \infty$, w.p.1, where the null set does not depend on a. Since $L(\vartheta_0)$ is positive definite and $a \neq 0$, $a^\top L(\vartheta_0)a > 0$. Thus, the Lyapunov condition is fulfilled if we can show that w.p.1

$$\frac{1}{n^{1+\nu/2}} \sum_{i=1}^{n} \mathbb{E}_n^*\Big(\big|a^\top l(X_i, Y_{i,n}^*, \vartheta_n)\big|^{2+\nu}\Big) \longrightarrow 0, \quad \text{as } n \to \infty.$$

Apply 6.5.2(vii) and choose $\nu = \delta$ to get

$$\frac{1}{n^{\delta/2}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_n^*\Big(\big|a^\top l(X_i, Y_{i,n}^*, \vartheta_n)\big|^{2+\delta}\Big) \leq \frac{\|a\|^{2+\delta}}{n^{\delta/2}} \frac{1}{n} \sum_{i=1}^{n} h_{l,e}(X_i, Y_i, \vartheta_n).$$

Corollary 5.67 together with assumption 6.5.1(iii) completes the proof. □

As we have outlined in the introduction to this chapter,

$$R_n^{1*}(u) = R_n^*(u) + S_n^*(u),$$

where $S_n^*(u)$ is defined in (6.22). The main part now is to handle the process S_n^*.

Note that assumptions 6.5.2(iv) and 6.5.1(iii) imply

$$\mathbb{P}_n^*(\|\vartheta_n^* - \vartheta_0\| > \varepsilon) \longrightarrow 0, \quad \text{as } n \to \infty, \text{ w.p.1}, \tag{6.25}$$

for $\varepsilon > 0$. Except for an $o_{\mathbb{P}_n^*}(1)$ term we can therefore assume that ϑ_n^* and ϑ_n are in the neighborhood V from assumption 6.5.2(ix) and we can apply Taylor's expansion to get

$$m(x, \vartheta_n^*) = m(x, \vartheta_n) + (\vartheta_n^* - \vartheta_n)^\top w(x, \hat{\vartheta}_n^*(x)),$$

where $\hat{\vartheta}_n^*(x)$ is in the line segment connecting ϑ_n^* and ϑ_n. Now, as under (6.15), we set $W(t) = W(t, \vartheta_0) = (W_1(t, \vartheta_0), \ldots, W_{p+q}(t, \vartheta_0))^\top$, where

$$W_i(t) = W_i(t, \vartheta_0) = \mathbb{E}\left(w_i(X, \vartheta_0)\mathrm{I}_{\{\beta_0^\top X \le t\}}\right).$$

If we insert this in S_n^*, then the following decomposition is obtained.

$$\begin{aligned} S_n^*(u) &= n^{1/2}\big(\vartheta_n^* - \vartheta_n\big)^\top n^{-1}\sum_{i=1}^n w(X_i, \hat{\vartheta}_n^*(x))\mathrm{I}_{\{\beta_n^\top X_i \le u\}} + o_{\mathbb{P}_n^*}(1) \\ &= n^{1/2}\big(\vartheta_n^* - \vartheta_n\big)^\top W(u, \vartheta_0) \\ &\quad + n^{1/2}\big(\vartheta_n^* - \vartheta_n\big)^\top n^{-1}\sum_{i=1}^n \big(w(X_i, \hat{\vartheta}_n^*(X_i)) - w(X_i, \vartheta_0)\big)\mathrm{I}_{\{\beta_n^\top X_i \le u\}} \\ &\quad + n^{1/2}\big(\vartheta_n^* - \vartheta_n\big)^\top \Big(n^{-1}\sum_{i=1}^n w(X_i, \vartheta_0)\mathrm{I}_{\{\beta_n^\top X_i \le u\}} - W(u, \vartheta_0)\Big) \\ &\quad + o_{\mathbb{P}_n^*}(1). \end{aligned} \tag{6.26}$$

Lemma 6.31 *Let $\hat{\vartheta}_n^* : \mathbb{R}^p \to V$ be a measurable function such that $\hat{\vartheta}_n^*(x)$ lies for each $x \in \mathbb{R}^p$ in the line segment that connects ϑ_0^* and ϑ_0 and assume that 6.5.1(iii), 6.5.2(iv), 6.5.2(ix), 6.5.2(x), and 6.5.2(xi) hold. Then, w.p.1, for $1 \le j \le p+q$, as $n \to \infty$,*

(i) $\sup_{u\in\mathbb{R}} \left|n^{-1}\sum_{i=1}^n w_j(X_i, \vartheta_0)I_{\{\beta_n^\top X_i \le u\}} - W_j(u, \vartheta_0)\right| \longrightarrow 0,$

(ii) $\sup_{u\in\mathbb{R}} \left|n^{-1}\sum_{i=1}^n \big(w_j(X_i, \hat{\vartheta}_n^*(X_i)) - w_j(X_i, \vartheta_0)\big)I_{\{\beta_n^\top X_i \le u\}}\right| = o_{\mathbb{P}_n^*}(1).$

Proof Since $w_j(X, \vartheta_0)$ is integrable and the collection of half-spaces in $\mathbb{R}^p$ forms a GC class, we get from Kosorok (2008, Corollary 9.27) that w.p.1

$$\sup_{\beta\in\mathbb{R}^p,\, u\in\mathbb{R}} \left|n^{-1}\sum_{i=1}^n w_j(X_i, \vartheta_0)\mathrm{I}_{\{\beta^\top X_i \le u\}} - \mathbb{E}\big(w_j(X, \vartheta_0)\mathrm{I}_{\{\beta^\top X \le u\}}\big)\right| \longrightarrow 0, \quad \text{as } n \to \infty.$$

Therefore, we obtain from 6.5.1(iii) that for every $\varepsilon > 0$

$$\begin{aligned} &\limsup_{n\to\infty} \sup_{u\in\mathbb{R}} \left|n^{-1}\sum_{i=1}^n w_j(X_i, \vartheta_0)\mathrm{I}_{\{\beta_n^\top X_i \le u\}} - W_j(u, \vartheta_0)\right| \\ &\quad \le \limsup_{n\to\infty} \sup_{\beta\in\mathbb{R}^p,\, u\in\mathbb{R}} \left|n^{-1}\sum_{i=1}^n w_j(X_i, \vartheta_0)\mathrm{I}_{\{\beta^\top X_i \le u\}} - \mathbb{E}\big(w_j(X, \vartheta_0)\mathrm{I}_{\{\beta^\top X \le u\}}\big)\right| \\ &\qquad + \sup_{\|\beta-\beta_0\|<\varepsilon,\, u\in\mathbb{R}} \left|\mathbb{E}\big(w_j(X, \vartheta_0)\mathrm{I}_{\{\beta^\top X \le u\}}\big) - W_j(u, \vartheta_0)\right| \\ &\quad = \sup_{\|\beta-\beta_0\|<\varepsilon,\, u\in\mathbb{R}} \left|\mathbb{E}_n\big(w_j(X, \vartheta_0)\mathrm{I}_{\{\beta^\top X \le u\}}\big) - W_j(u, \vartheta_0)\right|, \end{aligned}$$

w.p.1. But the last term on the right side tends to 0, as $\varepsilon \to 0$, under the stated assumptions, with similar arguments as we used in the proof of Lemma 6.26. For the second part, we get from (6.25) that

$$\begin{aligned}\sup_{u\in\mathbb{R}}\Big|n^{-1}\sum_{i=1}^{n}\big(w_j(X_i,\hat{\vartheta}_n^*(X_i)) - w_j(X_i,\vartheta_0)\big)\mathrm{I}_{\{\beta_n^\top X_i\le u\}}\Big| \\ \le n^{-1}\sum_{i=1}^{n}\sup_{\|\vartheta-\vartheta_0\|<\varepsilon}\big|w_j(X_i,\vartheta)-w_j(X_i,\vartheta_0)\big| + o_{\mathbb{P}_n^*}(1).\end{aligned}$$

Now, by condition 6.5.2(x),

$$\begin{aligned}n^{-1}\sum_{i=1}^{n}\sup_{\|\vartheta-\vartheta_0\|<\varepsilon}\big|w_j(X_i,\vartheta)-w_j(X_i,\vartheta_0)\big| \\ \longrightarrow \mathbb{E}\Big(\sup_{\|\vartheta-\vartheta_0\|<\varepsilon}\big|w_j(X_i,\vartheta)-w_j(X_i,\vartheta_0)\big|\Big),\end{aligned}$$

as $n \to \infty$, w.p.1. Due to the assumptions 6.5.2(ix) and 6.5.2(x), the last expectation tends to 0, as $\varepsilon \to 0$, by an application of the dominated convergence theorem. This proves the lemma. □

Under the assumptions of Lemma 6.31 we get from (6.26) that uniformly in u

$$S_n^*(u) = n^{1/2}\big(\vartheta_n^* - \vartheta_n\big)^\top W(u,\vartheta_0) + o_{\mathbb{P}_n^*}(1), \quad \text{w.p.1.}$$

Now, use the asymptotic linear representation of ϑ_n^* of condition 6.5.2(vi) for further modification of S_n^* to get

$$S_n^*(u) = n^{-1/2}\sum_{i=1}^{n} l^\top(X_i, Y_{i,n}^*, \vartheta_n)W(u) + o_{\mathbb{P}_n^*}(1), \quad \text{w.p.1}, \tag{6.27}$$

uniformly in u. Here and in the rest of this section we use $W(u)$ for $W(u,\vartheta_0)$.

Theorem 6.32 *Assume that conditions 6.5.1(i)–6.5.1(iii) and 6.5.2(v)–6.5.2(xii) hold. Then, w.p.1, R_n^{1*} converges in $D[-\infty,\infty]$ to a centered Gaussian process $\bar{R}_\infty^1$ with covariance function*

$$\begin{aligned}\bar{K}^1(s,t) = K(s,t) + W^\top(s)L(\vartheta_0)W(t) \\ - W^\top(s)\mathbb{E}(I_{\{\beta_0^\top X\le t\}}(Y - m(\beta_0^\top X,\theta_0))l(X,Y,\vartheta_0)) \\ - W^\top(t)\mathbb{E}(I_{\{\beta_0^\top X\le s\}}(Y - m(\beta_0^\top X,\theta_0))l(X,Y,\vartheta_0)),\end{aligned}$$

where $K(s,t)$ is the covariance function of the BMEP given in Theorem 6.27.

Proof Due to the assumed conditions, we can use the representation of S_n^* obtained under (6.27) to get that the two sequences $(R_n^{1*})_{n\in\mathbb{N}}$ and $(\hat{R}_n^{1*})_{n\in\mathbb{N}}$, where

$$\hat{R}_n^{1*}(u) = n^{-1/2}\sum_{i=1}^{n}\Big(\big(Y_{i,n}^* - m(\beta_n^\top X_i, \theta_n)\big)\mathrm{I}_{\{\beta_n^\top X_i \le u\}} - l^\top(X_i, Y_{i,n}^*, \vartheta_n)W(u)\Big),$$

are asymptotically equivalent in the sense of Billingsley (1968, Theorem 4.1). To prove the assertion, we apply Billingsley (1968, Theorem 15.6) to $(\hat{R}_n^{1*})_{n\in\mathbb{N}}$. Note that

$$\hat{R}_n^{1*}(u) = R_n^*(u) - n^{-1/2}\sum_{i=1}^{n} l^\top(X_i, Y_{i,n}^*, \vartheta_n)W(u)$$

and $(R_n^*)_{n\in\mathbb{N}}$ is tight in $D[-\infty,\infty]$ according to Theorem 6.27. Furthermore, the proof of Lemma 6.30 shows that $n^{-1/2}\sum_{i=1}^{n} l^\top(X_i, Y_{i,n}^*, \vartheta_n)$ converges to a zero mean multivariate normal distribution with covariance matrix $L(\vartheta_0)$, w.p.1. By assumption 6.5.2(xi), $W(\cdot)$ is continuous. Thus, $n^{-1/2}\sum_{i=1}^{n} l^\top(X_i, Y_{i,n}^*, \vartheta_n)W(u)$ is tight in $C[-\infty,\infty]$ and therefore also tight in $D[-\infty,\infty]$. All in all, the tightness of $(\hat{R}_n^{1*})_{n\in\mathbb{N}}$ results, w.p.1.

It remains to show that the fidis of $\hat{R}_n^{1*}$ converge in distribution to those of $\bar{R}_\infty^1$. For this, let $k\in\mathbb{N}$, take $-\infty \le u_1 < \ldots < u_k \le \infty$, and $0 \ne a \in \mathbb{R}^k$. To apply Cramér-Wold, we have to show that w.p.1

$$Z_n^* = \sum_{j=1}^{k} a_j \hat{R}_n^{1*}(u_j) \longrightarrow \mathcal{N}(0, a^\top \Sigma a), \quad \text{for } n\to\infty,$$

in distribution, where $\Sigma = (\sigma_{r,s})_{1\le r,s\le k}$ and $\sigma_{r,s} = \mathrm{COV}(\bar{R}_\infty^1(u_r), \bar{R}_\infty^1(u_s)) = \bar{K}^1(u_r, u_s)$. A simple rearrangement of the terms in Z_n^* results in

$$\begin{aligned} Z_n^* &= \sum_{i=1}^{n} \frac{Y_{i,n}^* - m(X_i, \vartheta_n)}{\sqrt{n}} \sum_{j=1}^{k} a_j \mathrm{I}_{\{\beta_n^\top X_i \le u_j\}} - \frac{l^\top(X_i, Y_{i,n}^*, \vartheta_n)}{\sqrt{n}} \sum_{j=1}^{k} a_j W(u_j) \\ &= \sum_{i=1}^{n} \xi_{i,n}^* A_{i,n} - \eta_{i,n}^{*\top} B, \end{aligned}$$

where $\xi_{i,n}^* = n^{-1/2}(Y_{i,n}^* - m(X_i, \vartheta_n))$ and $\eta_{i,n}^* = n^{-1/2} l(X_i, Y_{i,n}^*, \vartheta_n)$. These variables are centered and $(\xi_{1,n}^*, \eta_{1,n}^*), \ldots, (\xi_{n,n}^*, \eta_{n,n}^*)$ are independent. Furthermore, $A_{i,n}$ and B are deterministic with respect to $\mathbb{P}_n^*$. For the variance of Z_n^*, this results in

$$\mathrm{VAR}_n^*(Z_n^*) = \sum_{i=1}^n A_{i,n}^2 \mathrm{VAR}_n^*(\xi_{i,n}^*) + \sum_{i=1}^n B^\top \mathbb{E}_n^*\big(\eta_{i,n}^* \eta_{i,n}^{*\top}\big) B$$
$$- 2B^\top \sum_{i=1}^n \mathbb{E}_n^*(\xi_{i,n}^* \eta_{i,n}^*) A_{i,n}.$$

As we have seen in the proof of Theorem 6.27,

$$\sum_{i=1}^n A_{i,n}^2 \mathrm{VAR}_n^*(\xi_{i,n}^*) \longrightarrow \sum_{1\le r,s\le k} a_r K(u_s, u_r) a_s, \quad \text{as } n \to \infty, \text{ w.p.1}.$$

Assumption 6.5.2(viii) guarantees that w.p.1, as $n \to \infty$,

$$\sum_{i=1}^n B^\top \mathbb{E}_n^*\big(\eta_{i,n}^* \eta_{i,n}^{*\top}\big) B \longrightarrow B^\top L(\vartheta_0) B = \sum_{1\le r,s\le k} a_s W^\top(u_s) L(\theta_0) W(u_r) a_r.$$

For the last term, conditions 6.5.1(iii), 6.5.2(xii) together with Lemma 6.26 imply that, w.p.1, as $n \to \infty$,

$$B^\top \sum_{i=1}^n \mathbb{E}_n^*(\xi_{i,n}^* \eta_{i,n}^*) A_{i,n}$$
$$= \sum_{1\le r,s\le k} a_r W^\top(u_r) \frac{1}{n} \sum_{i=1}^n h_{cov}(X_i, Y_i, \vartheta_n) \mathrm{I}_{\{\beta_n^\top X_i \le u_s\}} a_s$$
$$\longrightarrow \sum_{1\le r,s\le k} a_r W^\top(u_r) \mathbb{E}\big((Y - m(\beta_0^\top X, \theta_0)) l(X, Y, \vartheta_0) \mathrm{I}_{\{\beta_0^\top X \le u_s\}}\big) a_s.$$

All in all this shows that w.p.1, as $n \to \infty$

$$\mathrm{VAR}_n^*(Z_n^*) \longrightarrow \sum_{1\le r,s\le k} a_r \bar{K}^1(u_r, u_s) a_s = a^\top \Sigma a.$$

If $a^\top \Sigma a = 0$, Chebyshev's inequality implies that $Z_n^* = o_{\mathbb{P}_n^*}(1)$ and we have that

$$Z_n^* \longrightarrow \mathcal{N}(0, a^\top \Sigma a), \quad \text{as } n \to \infty.$$

Now assume that $a^\top \Sigma a > 0$. According to Serfling (1980, Corollary to Theorem 1.9.3), the validity of the Lyapunov condition

$$\frac{1}{\mathrm{VAR}_n^*(Z_n^*)^{(2+\nu)/2}} \sum_{i=1}^n \mathbb{E}_n^*\big(|\xi_{i,n}^* A_{i,n} - B^\top \eta_{i,n}^*|^{2+\nu}\big) \longrightarrow 0, \quad \text{as } n \to \infty, \text{ w.p.1},$$

for some $\nu > 0$, implies the asymptotic normality of Z_n^*. Note that

$$\begin{aligned}
&\mathbb{E}_n^*\big(|\xi_{i,n}^* A_{i,n} - B^\top \eta_{i,n}^*|^{2+\nu}\big) \\
&\quad = \sum_{i=1}^n \Big(\mathbb{E}_n^*\big(|\xi_{i,n}^* A_{i,n} - B^\top \eta_{i,n}^*|^{2+\nu}\big)^{1/(2+\nu)}\Big)^{2+\nu} \\
&\quad \le \sum_{i=1}^n \Big(\mathbb{E}_n^*\big(|\xi_{i,n}^* A_{i,n}|^{2+\nu}\big)^{1/(2+\nu)} + \mathbb{E}_n^*\big(|B^\top \eta_{i,n}^*|^{2+\nu}\big)^{1/(2+\nu)}\Big)^{2+\nu} \\
&\quad \le 2^{2+\nu} \sum_{i=1}^n \max\Big(\mathbb{E}_n^*\big(|\xi_{i,n}^* A_{i,n}|^{2+\nu}\big), \mathbb{E}_n^*\big(|B^\top \eta_{i,n}^*|^{2+\nu}\big)\Big) \\
&\quad \le 2^{2+\nu}\Big(\sum_{i=1}^n \mathbb{E}_n^*\big(|\xi_{i,n}^* A_{i,n}|^{2+\nu}\big) + \sum_{i=1}^n \mathbb{E}_n^*\big(|B^\top \eta_{i,n}^*|^{2+\nu}\big)\Big).
\end{aligned}$$

Conditions 6.5.1(i) and 6.5.2(vii) allow us to apply the same arguments used to verify the Lyapunov condition in the proof of Theorem 6.27. This completes the proof. □

Remark 6.33 Note that $\bar{K}^1$ matches the covariance given under (6.19). Thus, the bootstrap version of the EMEP converges to the same process as the EMEPE does.

6.6 Exercises

Exercise 6.1 The Kolmogorov-Smirnov (D_n) and Cramér-von Mises (W_n^2) statistics are used in the GOF test. If you want to use another statistics, which property is necessary so that all GOF-related theorems hold true.

Exercise 6.2 At which point in the mathematical framework of the marked empirical process is the fact necessary that at least one of the covariates is a continuous random variable.

Exercise 6.3 Compare the performance of the GOF test using the Kolmogorov-Smirnov and Cramér-von Mises statistics. For instance, extend the plots shown in Fig. 6.18.

Exercise 6.4 Plot the $p-$values of the GOF test based on the Kolmogorov-Smirnov (KS) statistics against the $p-$values of the GOF test based on the Cramér-von Mises (CvM) statistics. Make sure that each $p-$value pair was generated on the same original and bootstrap datasets.

Investigate a situation where the p-value based on the KS is small and the p-value based on CvM is large and vice versa. Can you modify one of the datasets in order to make the difference between the p-values even larger?

References

Billingsley P (1968) Convergence of probability measures. Wiley, New York

Billingsley P (1995) Probability and measure, 3rd edn. Wiley Series in Probability and Mathematical Statistics. Wiley, New York

Dikta G, Kvesic M, Schmidt C (2006) Bootstrap approximations in model checks for binary data. J Am Stat Assoc 101(474):521–530

Khmaladze EV (1982) Martingale approach in the theory of goodness-of-fit tests. Theory of Probability & Its Applications 26(2):240–257

Kosorok M (2008) Introduction to empirical processes and semiparametric inference. Springer series in statistics. Springer, New York

Nikabadze A, Stute W (1997) Model checks under random censorship. Stat Probab Lett 32(3):249–259

Serfling RJ (1980) Approximation theorems of mathematical statistics. Wiley Series in Probability and Mathematical Statistics. Wiley, New York

Shorack GR (2000) Probability for statisticians. Springer Texts in Statistics. Springer, New York

Shorack GR, Wellner JA (1986) Empirical processes with applications to statistics. Wiley Series in Probability and Mathematical Statistics. Wiley, New York

Stute W (1997) Nonparametric model checks for regression. Ann Stat 25(2):613–641

Stute W, Zhu LX (2002) Model checks for generalized linear models. Scandinavian Journal of Statistics Theory and Applications 29(3):535–545

Stute W, González Manteiga W, Presedo Quindimil M (1998) Bootstrap approximations in model checks for regression. J Am Stat Assoc 93(441):141–149

van der Vaart AW, Wellner JA (1996) Weak Convergence and Empirical Processes: With Applications to Statistics. Springer Series in Statistics, Springer, New York

van Heel M, Dikta G, Braekers R (2019) Bootstrap based goodness-of-fit tests for binary multivariate regression models. Technical report, Fachhochschule Aachen