Chapter 9
Additional Considerations and Final Thoughts

The statistical study of DTRs and associated methods of estimation is a young and growing field. As such, there are many topics which are only beginning to be explored. In this chapter, we point to some new developments and areas of research in the field.

9.1 Variable Selection

In estimating optimal adaptive treatment strategies, the variables used to tailor treatment to patient characteristics are typically hand-picked by experts who seek to use a minimum set of variables routinely available in clinical practice. However, studies often use a large set of easy-to-measure covariates (e.g., multiple surveys of mental health status and functioning) from which a smaller subset of variables must be selected for any practical implementation of treatment tailoring. It may therefore be desirable to be able to select tailoring variables with which to index the class of regimes using automated or data-adaptive procedures. It has been noted that prediction methods such as boosting could aid in selecting variables to adapt treatments (LeBlanc and Kooperberg 2010); many such methods can be applied with ease, particularly to the regression-based approaches to estimating optimal DTRs, however their ability to select variables for strong *interactions* with treatment rather than simply strong predictive power may require special care and further study.

Recall the distinction between *predictive* variables (used to increase precision of estimates) and *prescriptive* variables (used to adapt treatment strategies to patients), i.e. tailoring variables (Gunter et al. 2007). In the Q-learning notation, predictive variables correspond to the H_{j0} terms in the Q-function associated with parameters β, while the prescriptive or tailoring variables are those contained in H_{j1}, associated with parameters ψ. Tailoring variables must *qualitatively* interact with the treatment, meaning that the choice of optimal treatment varies for different values of such variables. The usefulness of a prescriptive variable can be characterized by

B. Chakraborty and E.E.M. Moodie, *Statistical Methods for Dynamic Treatment Regimes*, Statistics for Biology and Health 76, DOI 10.1007/978-1-4614-7428-9_9,

the magnitude of the interaction and the proportion of the population for whom the optimal action changes given the knowledge of the variable (Gunter et al. 2007).

We will focus the discussion in this section on the randomized trial setting, so that variable selection is strictly for the purposes of optimal treatment tailoring, rather than elimination of bias due to confounding. Further, we will restrict attention to the one-stage setting, as to date there have been no studies on the use of variable selection for dynamic treatment regimes in the multi-stage setting.

9.1.1 Penalized Regression

Lu et al. (2013) proposed an adaptation of the lasso which penalizes only interaction terms. Specifically, they consider the loss function

$$L_n(\psi, \beta, \alpha) = \mathbb{P}_n[Y_i - \phi(O_i; \beta) - \psi^T O_i(A_i - \pi(O_i))]^2 \tag{9.1}$$

where the covariate vector O_i is augmented by a column of 1s and has total length $p+1$, $\pi(o) = P(A = 1|O = o; \alpha)$ is the propensity score for a binary treatment A and $\phi(O)$ is an arbitrary function. Lu et al. (2013) noted that the estimating function found by taking the derivative of the loss function $L_n(\psi, \beta, \alpha)$ with respect to ψ corresponds to an A-learning method of estimation, and is therefore robust to misspecification of the conditional mean model $\phi(O; \beta)$ for the response Y in the sense that the estimator requires correct specification of either the propensity score or the mean model $\phi(O; \beta)$. The decision (or treatment interaction) parameters ψ are then shrunk using an adaptive lasso which penalizes parameters with a weight inversely proportional to their estimated value, solving

$$\min_{\psi} L_n(\psi, \hat{\beta}, \hat{\alpha}) + \lambda_n \sum_{j=1}^{p+1} |\hat{\psi}_j|^{-1} |\psi_j|$$

where $\hat{\psi}, \hat{\beta}$ are solutions to Eq. (9.1), $\hat{\alpha}$ is a consistent estimate of the propensity score model parameters, and λ_n is a tuning parameter that may be selected using cross-validation or some form of Bayesian Information Criterion (BIC). By penalizing the interaction parameters with the inverse of their estimated values, important interactions (i.e. those estimated to have large coefficients) will receive little penalty, while those with small estimates will be highly penalized.

Lu et al. (2013) showed that under standard regularity conditions, the estimators of the parameters ψ resulting from the penalized regression will be asymptotically normal and the set of selected treatment-covariate interactions will equal the set of treatment-interaction which are truly non-zero. The properties of the penalized estimator in multi-stage or non-regular settings were not examined. The estimator was compared to the unpenalized estimator $\hat{\psi}$ that results from solving Eq. (9.1) in low and high dimensional problems (10 and 50 variables, respectively). Using a linear working model for $\phi(O_i; \beta)$, the penalized estimator selected the truly non-zero

interaction terms with very high probability in samples of size 100 or larger. In high dimensional settings, the penalized estimator increased the selection of the correct treatment choice relative to the unpenalized estimator by 7–8 %; in low dimensional settings, the improvement was more modest (2–3 %).

9.1.2 Variable Ranking by Qualitative Interactions

As proposed by Gunter et al. (2007, 2011b), the *S-score* for a (univariate) variable O is defined as:

$$S_O = \mathbb{P}_n \left\{ \max_{a \in \mathscr{A}} \mathbb{P}_n [Y|A = a, O] - \max_{a \in \mathscr{A}} \mathbb{P}_n [Y|A = a] \right\}.$$

The S-score of a variable O captures the expected increase in the response that is observed by adapting treatment based on the value of that variable. S-scores combine two characteristics of a useful tailoring variable: the interaction of the variable with the treatment and the proportion of the population exhibiting variability in its value. A high S-score for a variable is indicative of a strong qualitative interaction between the variable and the treatment, as well as a high proportion of patients for whom the optimal action would change if the value of the variable were taken into consideration. Thus, S-scores may be used to rank variables and select those that have the highest scores. The performance of the S-score ranking method was found to be superior to the standard lasso (Tibshirani 1996) in terms of consistent selection of a small number of variables from a large set of covariates of interest.

In the real-data implementation of the S-score ranking performed by Gunter et al. (2007), each variable was evaluated separately, without taking into account potential correlation between variables. Two variables that are highly correlated may have similar S-scores (Biernot and Moodie 2010) but may not both be necessary for decision making. The S-score may be modified in a straight-forward fashion to examine the usefulness of sets of variables, O' given the use of others, O, by considering, for example,

$$S_{O'|O} = \mathbb{P}_n \left\{ \max_{a \in \mathscr{A}} \mathbb{P}_n \left[Y|A = a, O, O' \right] - \max_{a \in \mathscr{A}} \mathbb{P}_n [Y|A = a, O] \right\}.$$

Thus, the S-score approach could be used to select the variable, O, with the highest score, then select a second variable, O', with the highest S-score given the use of O as a prescriptive variable, and so on.

For $i = 1, \ldots, n$ subjects and $j = 1, \ldots, p$ possible tailoring variables, Gunter et al. (2007, 2011b) proposed an alternative score, also based on both the strength of interaction as measured by

$$D_j = \max_{1\le i\le n} \left(\mathbb{P}_n[Y|O_j = o_{ij}, A = 1] - \mathbb{P}_n[Y|O_j = o_{ij}, A = 0]\right) - \min_{1\le i\le n} \left(\mathbb{P}_n[Y|O_j = o_{ij}, A = 1] - \mathbb{P}_n[Y|O_j = o_{ij}, A = 0]\right)$$

and the proportion of the population for whom the optimal decision differs if a variable is used for tailoring, captured by

$$P_j = \mathbb{P}_n \mathbb{I}\left[\operatorname*{argmax}_a \mathbb{P}_n[Y|O_j = o_{ij}, A = a] \neq a^*\right]$$

where $a^* = \operatorname{argmax}_a \mathbb{P}_n[Y|A = a]$ is the optimal decision in the absence of tailoring. These values are combined to form another means of ranking the importance of tailoring variables, called the U-score:

$$U_j = \left(\frac{D_j - \min_{1\le k\le p} D_k}{\max_{1\le k\le p} D_k - \min_{1\le k\le p} D_k}\right)\left(\frac{P_j - \min_{1\le k\le p} P_k}{\max_{1\le k\le p} P_k - \min_{1\le k\le p} P_k}\right).$$

Gunter et al. (2007, 2011b) suggested the use of the S- and U-scores in combination with lasso:

1. Select variables that are predictive of the outcome Y from among the variables in (H_{10}, AH_{11}), using cross-validation or the BIC to select the penalty parameter.
2. Rank each variable O_j using the S- or U-score, retaining the predictive variables selected in step (1) to reduce the variability in the estimated mean response. Choose the M most highly-ranked variables, where M is the cardinality of the variables in H_{11} for which the S- or U-score is non-zero.
3. Create nested subsets of variables.
   (a) Let H_{11}^* be the top M variables found in step (2), and let H_{10}^* denote the union of the predictive variables chosen in step (1) and H_{11}^*. Let M^* denote the cardinality of (H_{10}^*, H_{11}^*).
   (b) Run a weighted lasso where all main effect and interaction variables chosen in step (1) only have weight 1, and all interaction variables chosen in step (2) are given a weight $0 < w \le 1$ which is a non-decreasing function of the U- or S-score. This downweights the importance of the prescriptive variables, which are favored by lasso.
   (c) Create M^* nested subsets based on the order of entry of the M^* variables in the weighted lasso.
4. Choose from among the subsets based on the highest expected response, or alternatively, the highest adjusted gain in the outcome relative to not using any tailoring variables.

The variable selection approaches based on the S- and U-scores were found to perform well in simulation, leading to variable choices that provided higher expected outcomes than lasso alone (Gunter et al. 2007, 2011b).

9.1.3 Stepwise Selection

Gunter et al. (2011a) suggested that the qualitative ranking of the previous section is complex and difficult to interpret, and instead proposed the use of a stepwise procedure, using the expected response conditional on treatment A and covariates O, as the criterion on which to select or omit tailoring variables.

The suggested approach begins by fitting a regression model for the response Y as a function of treatment only, and estimating the mean response to the overall ("un-tailored") optimal treatment; denote this by $\hat{V}_0^*$. Next, let $\mathscr{C}$ contain the treatment variable as well as all variables which are known to be important predictors of the response. Fit a regression model for the response Y as a function of treatment and all variables in $\mathscr{C}$ and estimate the mean response to the overall (un-tailored) optimal treatment when the predictors in $\mathscr{C}$ are included in the model; denote this by $\hat{V}_C^*$. A key quantity that will be used to decide variable inclusion or exclusion is the *adjusted value* of the model. For $\mathscr{C}$, the adjusted value is $AV_C = (\hat{V}_C^* - \hat{V}_0^*)/|\mathscr{C}|$ where $|\mathscr{C}|$ is the rank of the model matrix used in the estimation of the response conditional on the variables in $\mathscr{C}$.

Letting $\mathscr{E}$ denote all eligible variables, both predictive variables and treatment-covariate interaction terms, not included in $\mathscr{C}$. The procedure is then carried out by performing forward selection and backwards elimination at each step.

Forward selection: For each variable $e \in \mathscr{E}$,

1. Estimate the predictive model using all the variables in $\mathscr{C}$ plus the variable e.
2. Optimize the estimated predictive model over the treatment actions to obtain the optimal mean response, $\hat{V}_E^*$, and calculate the adjusted value, $AV_e = (\hat{V}_e^* - \hat{V}_0^*)/|\mathscr{C} + e|$.
3. Retain the covariate e^* which results in the largest value of AV_e.

Backward elimination: For each variable $c \in \mathscr{C}$,

1. Estimate the predictive model using all the variables in $\mathscr{C}$ except the variable c.
2. Optimize the estimated predictive model over the treatment actions to obtain the optimal mean response, $\hat{V}_{-c}^*$, and calculate the adjusted value, $AV_{-c} = (\hat{V}_{-c}^* - \hat{V}_0^*)/|\mathscr{C} - c|$.
3. Let c^* be the covariate which results in the largest value of AV_{-c}.

If each of AV_C, AV_{e^*}, and AV_{-c^*} are negative, the stepwise procedure is complete and no further variable selection is required. If $AV_{e^*} > \max\{AV_C, AV_{-c^*}\}$, e^* is included in $\mathscr{C}$ and AV_C is set to AV_{e^*}; otherwise, if $AV_{-c^*} > \max\{AV_C, AV_{e^*}\}$, remove c from $\mathscr{C}$ and AV_C is set to AV_{-c^*}. Gunter et al. (2011a) suggested that all covariate main effects should be retained in a model in which a treatment-covariate interaction is present, and to group covariates relating to a single characteristic (e.g. dummy variables indicating covariate level for categorical variables).

In simulation, the stepwise method was found to have higher specificity but lower sensitivity than the qualitative interaction ranking approach of the previous section (Gunter et al. 2011a). That is, the stepwise procedure was less likely to falsely include variables which did not qualitatively interact with treatment, at the cost of

being less able to identify variables which did. However, the stepwise procedure is rather easier to implement and can be applied to different outcome types such as binary or count data.

Gunter et al. (2011c) used a similar, but more complex, method to perform variable selection while controlling the number of falsely significant findings by using bootstrap sampling and permutation thresholding in combination. The bootstrap procedure is used as a form of voting algorithm to ensure selection of variables that modify treatment in a single direction, while the permutation algorithm is used to maintain a family-wise error rate across the tests of significance for the coefficients associated with the tailoring variables.

9.2 Model Checking via Residual Diagnostics

There has been relatively little work on the topic of model checking for estimating optimal DTRs. The regret-regression approach of Henderson et al. (2010) is one of the first in which the issues of model checking and diagnostics were specifically addressed. Because regret-regression uses ordinary least squares for estimation of the model parameters, standard tools for regression model checking and diagnostics can be employed. In particular, Henderson et al. (2010) showed that residual plots can be used to diagnose model mis-specification. In fact, these standard approaches can and should be used whenever a regression-based approach to estimating DTR parameters, such as Q-learning or A-learning as implemented by Almirall et al. (2010), is taken.

Consider the following small example using Q-learning: data are generated such that $O_{11} \sim N(0,1)$ and $O_{21} \sim N(-0.5_0 + 0.5O_{11}, 1)$, treatment is randomly assigned at each stage with probability 1/2, and the binary tailoring variables are generated via

$$\begin{aligned} P[O_{12} = 1] &= P[O_{12} = -1] = 1/2, \\ P[O_{22} = 1|O_{12}, A_1] &= 1 - P[O_{22} = -1|O_{12}, A_1] = \text{expit}(0.1O_{12} + 0.1A_1). \end{aligned}$$

Thus the state variables are $O_1 = (O_{11}, O_{12})$ and $O_2 = (O_{21}, O_{22})$. Then for $\varepsilon \sim N(0,1)$,

$$Y = 0.5O_{11} - 0.5A_1 + 0.5O_{12}A_1 + 0.5O_{21} + A_2 + 1.4O_{22}A_2 + A_1A_2 + \varepsilon.$$

We fit three models. The first is correctly specified, the second omits the single predictive variable, O_{j1}, from the model for the Q-function at each stage, and the third omits the interaction A_jO_{j2} from the Q-function model. As observed in Fig. 9.1, residuals from the OLS fit at each stage of the Q-learning algorithm can be used to detect the omission of important predictors of the response, but may not be sufficiently sensitive to detect the omission of important tailoring variables from the Q-function model.

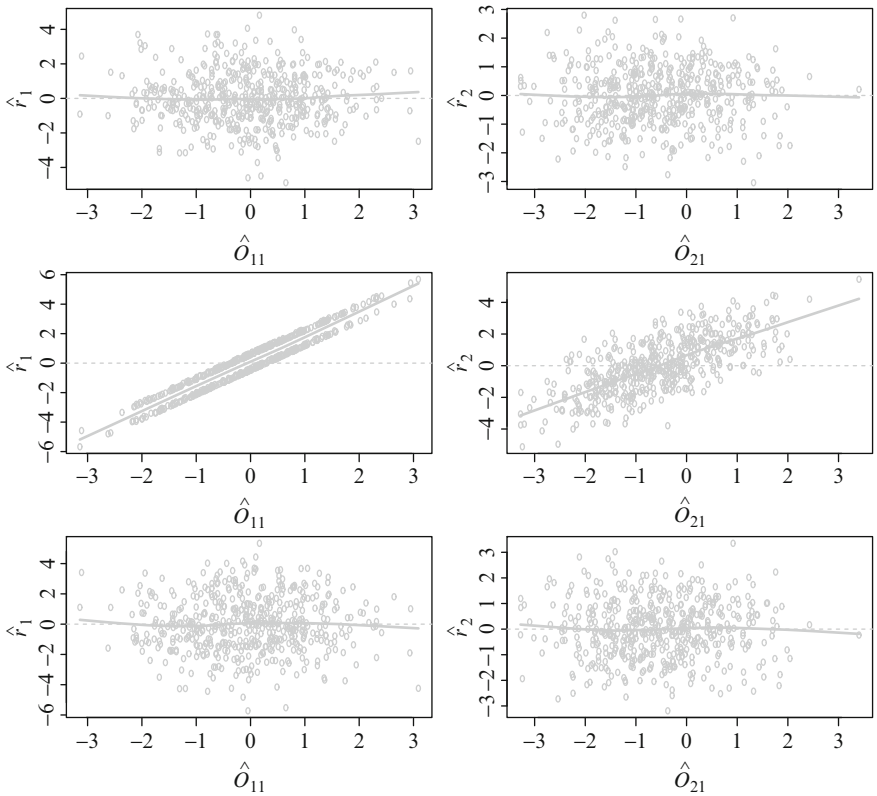

Fig. 9.1 Residual diagnostic plots for Q-learning using a simulated data set with n = 500. The first and second columns show plots for residuals at the first and second stages, respectively. The first row corresponds to a correctly specified Q-function model. In the second and third rows, Q-function models at each stage are mis-specified by the omission, respectively, of a predictive variable and an interaction with a tailoring variable

It is also possible to generalize the ideas of model-checking in regression to G-estimation, producing a type of residual that can be used to construct residual diagnostic plots. Recall that doubly-robust G-estimation can be based on the estimating function:

$$U = \sum_{j=0}^{K-1} U_j = \sum_{j=0}^{K-1} \left\{ G_{\text{mod},j}(\psi) - E[G_{\text{mod},j}(\psi)|H_j] \right\} \left\{ S_j(A_j) - E[S_j(A_j)|H_j] \right\}$$

where

$$G_{\text{mod},j}(\psi) \equiv G_{\text{mod},j}(\psi)(H_K, A_K, \psi_j) = Y - \gamma_j(H_j, A_j; \psi_j) + \sum_{m=j+1}^{K-1} \mu_m(H_m, A_m; \psi_m),$$

for $S_j(A_j)$ an analyst-specified function of H_j and A_j. In general, due to high dimensionality of the covariate space, estimation is made more tractable when parametric models are specified for:

1. The blip function $\gamma_j(h_j, a_j; \psi_j)$;
2. The expected potential outcome $E[G_{\text{mod},j}(\psi)|H_j; \varsigma_j]$;
3. The treatment model $E[A_j|H_j; \alpha_j]$, required for $E[S_j(A_j)|H_j; \alpha_j]$.

The first of these provides estimates of the decision rule parameters while the other two are considered nuisance models. Although some model mis-specification is permitted in the doubly-robust framework, there are efficiency gains when both models (2) and (3) are correct (Moodie et al. 2007).

Rich et al. (2010) note that, letting G_{ij} be $G_{\text{mod},j}(\psi)$ for subject i at stage j,

$$\begin{aligned} G_{ij} - E[G_{ij}|H_j; \varsigma_j(\psi_j)] & \\ &= \left\{ Y_i - \gamma_j(H_j, A_j; \psi_j)) + \sum_{m=j+1}^{K-1} \mu_m(H_m, A_m; \psi_m) \right\} - E[G_{ij}|H_j; \varsigma_j(\psi_j)] \\ &= Y_i - \left\{ E[G_{ij}|H_j; \varsigma_j(\psi_j)] - \sum_{m=j+1}^{K-1} \mu_m(H_m, A_m; \psi_m) + \gamma_j(H_j, A_j; \psi_j)) \right\} \end{aligned}$$

has mean zero conditional on history H_j, so that a fitted value for Y_i is given by

$$\hat{Y}_{ij}(\psi) = \left\{ E[G_{ij}|H_j; \varsigma_j(\psi_j)] - \sum_{m=j+1}^{K-1} \mu_m(H_m, A_m; \psi_m) + \gamma_j(H_j, A_j; \psi_j)) \right\}.$$

The residual for the ith individual at the jth stage is then defined to be

$$r_{ij}(\psi) = Y_i - \left\{ E[G_{ij}|H_j; \varsigma_j(\psi_j)] - \sum_{m=j+1}^{K-1} \mu_m(H_m, A_m; \psi_m) + \gamma_j(H_j, A_j; \psi_j)) \right\}.$$

To use the residual for model-checking purposes, estimates $\hat{\psi}$ and $\hat{\varsigma}_j(\hat{\psi}_j)$ must be substituted for the unknown parameters. The residuals r_{ij} can be used to verify the models $E[G_{ij}|H_j; \varsigma_j(\psi_j)]$ and $\gamma(h_j, a_j; \psi_j)$, diagnosing underspecification (that is, the omission of a variable) and checking the assumptions regarding the functional form in which covariates were included in the models.

Rich et al. (2010) considered a two-stage simulation, and examined plots of the first- and second-stage residuals against covariates and fitted values. The residual plots were able to detect incorrectly-specified models in a variety of settings, and appeared able to distinguish at which stage the model was mis-specified. While patterns in residual plots provide a useful indicator of problems with model specification, they do not necessarily indicate in *which* model a problem occurs, i.e. whether the problem is in the specification of the blip function or the expected counterfactual model.

Consider the following example, where data are generated as follows:

$$O_1 \sim N(0, 140)$$
$$O_2 \sim N(50 + 1.25O_1, 120)$$

$A_j = 1$ with probability p_j and $A_j = -1$ with probability $1 - p_j$ for $j = 1, 2$

$$Y \sim N(300 + 1.6O_1 + 1.2O_2, 300) - \mu_1(H_1, A_1; \psi_1) - \mu_2(H_2, A_2; \psi_2)$$

where $p_1 = \text{expit}(0.1 - 0.003O_1)$, $p_2 = \text{expit}(0.5 - 0.004O_2)$, and the regret functions $\mu_1(O_j, A_1; \psi_1)$, $\mu_2(H_2, A_2; \psi_2)$ are based on the linear blip functions

$$\gamma_1(O_1, A_1; \psi_1) = (170 - 3.4O_1)\mathbb{I}[A_1 = 1]$$
$$\gamma_2(H_2, A_2; \psi_2) = (420 - 2.8O_2)\mathbb{I}[A_1 = 1].$$

In Fig. 9.2, we plot the residuals for four different models, three of which have mis-specified components, from a single data set of size 500. The first and second models mis-specified the form of $E[G_{ij}|H_j; \zeta_j(\psi_j)]$, the expected counterfactual model, at stage one and two, respectively. The third model correctly specified the expected counterfactual models, but omitted O_1 and O_2 from the blip models at both stages. The fourth model was correctly specified. In the first, second, and fourth rows, the stage(s) where no models are mis-specified provide residual plots with no systematic patterns. However, if the expected counterfactual model (row 1 and 2) or the blip models (row 3) are mis-specified at one or both stages, obvious trends appear in the residual plots. As noted by Rich et al. (2010), mis-specification of the expected counterfactual model and the blip function result in similar patterns in the residual plots; it is therefore not possible to determine which model is incorrect simply by inspection of residual plots.

9.3 Discussion and Concluding Remarks

In this book, we have attempted to provide an introduction to the key findings in the statistical literature of dynamic treatment regimes. In Chaps. 1 and 2, we introduced the motivation for seeking evidence-based decision rules for treating chronic conditions, and outlined the key features of the structures of longitudinal data which are used to make inference about optimal treatment policies: observational follow-up studies and sequential multiple-assignment randomized trials. In the third chapter, we delved more deeply into the mathematics of the decision making problem and the reinforcement learning perspective. We also introduced Q-learning in Chap. 3, which is increasingly finding favor in the scientific community for the ease with which it can be implemented. Chapter 4 presented several semi-parametric methods arising from the causal inference literature: G-estimation and

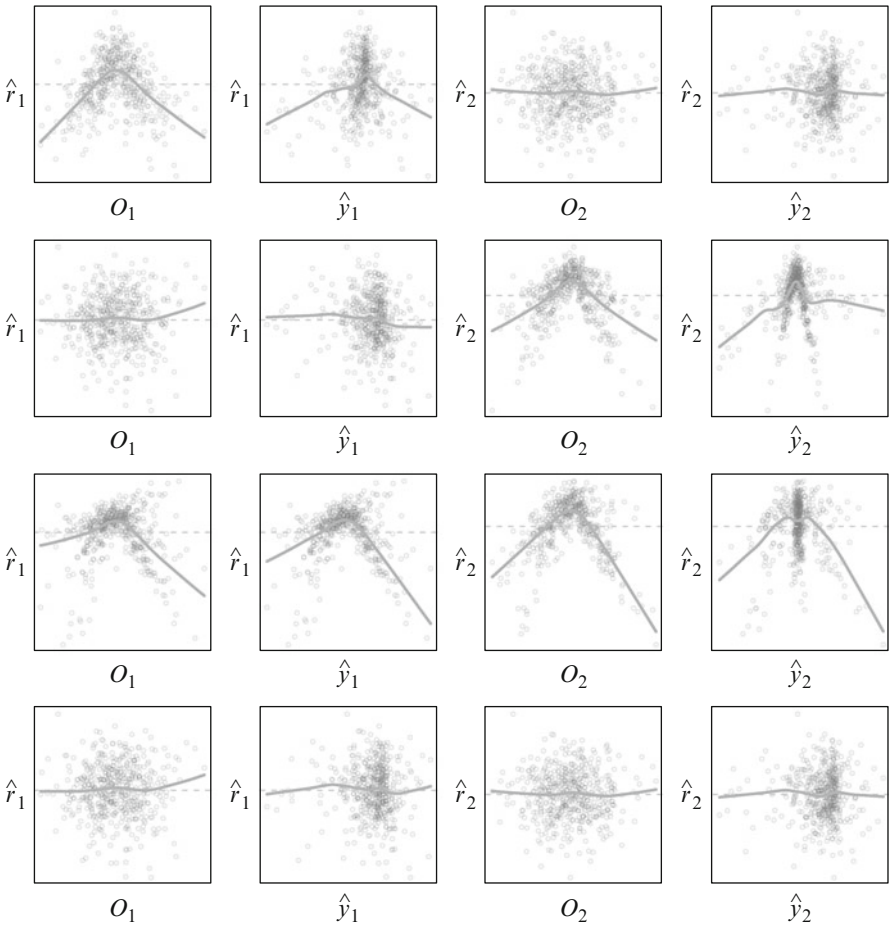

Fig. 9.2 Residual diagnostic plots for G-estimation using simulated data set with n = 500. The first two columns show plots for residuals and the first stage ($j = 1$), the last two for residuals at the second stage ($j = 2$). Specifically, the columns plot: (1) first stage residuals vs. O_1, (2) residuals vs. fitted values at the first stage, (3) second stage residuals vs. O_2, and (4) residuals vs. fitted values at the second stage. Rows correspond model choices: (1) $E[G_{\text{mod},1}(\psi)|O_1;\varsigma_1(\psi_1)]$ mis-specified, (2) $E[G_{\text{mod},2}(\psi)|H_2;\varsigma_2(\psi_2)]$ mis-specified, (3) $\gamma_1(O_1,A_1;\psi_1)$ and $\gamma_2(H_2,A_2;\psi_2)$ mis-specified, and (4) all models correctly specified. The solid grey curve indicates a loess smooth through the points

the regret-based methods including A-learning and regret-regression; where connections exist between methods, they were demonstrated. In Chap. 5, we turned our attention to methods that model regimes directly, including inverse probability weighting, marginal structural models, and classification-based approaches.

Our survey of estimation methods continued in Chap. 6, where the likelihood-based method of G-computation was demonstrated in both the frequentist and Bayesian contexts. In Chap. 7, we turned our attention to estimating DTRs for

alternative outcome types, including outcomes that are compound measures or multi-component in nature, as well as time-to-event and discrete valued. A range of methods have been applied in these settings, from Q-learning to marginal structural models to a likelihood-based approach.

Chapter 8 focused on improving estimation and inference, which presents a particular challenge in the dynamic treatment regime setting due to non-regularity of the estimators under certain underlying longitudinal data distributions, including those in which treatment has no effect. Three methods of bias reduction are considered: hard- and soft-thresholding, and penalized Q-learning. We then presented three bootstrap-based approaches to constructing confidence intervals which yield greatly improved coverage over any naively constructed interval at and near points in the parameter space that cause non-regularity of the estimators.

Finally, in this chapter, we have brought together a collection of topics that are at the forefront of research activity in dynamic regimes. The first two sections considered practical problems in implementing optimal DTR estimation: variable selection and model checking. In Sect. 9.1, we presented proposed approaches to the selection of tailoring variables, which differs from the usual problem of variable selection in that the analyst is seeking to find variables which qualitatively interact with treatment rather than those which are good predictors of the outcome. In the following section, we demonstrated the use of residual plots to assess model specification in Q-learning and G-estimation. As a summary of current or ongoing work, it is likely that this chapter is incomplete since the study of dynamic treatment regimes is, as a field, so active and is attracting new researchers from a diversity of backgrounds. The refinement and application of estimation techniques and the need to provide reliable measures of goodness-of-fit will continue to provide inspiration for many researchers in the coming years.

With the anticipated popularity of SMARTs in clinical and behavioral research, we foresee an inevitable complexity in the near future. Note that many interventions, either due to their very nature or due to logistical feasibility, need to be administered in group settings, requiring the design and analysis of cluster-randomized SMARTs. Some such complex trials are currently being considered. At the design level, cluster randomization would imply increased sample size requirements due to intra-class correlation, as expected. At the analysis level, on the other hand, it would open up several questions, e.g. how to incorporate random effects models or generalized estimating equations (GEE) methods into the framework of estimation techniques like Q-learning or G-estimation, whether the correlation would enhance the phenomenon of non-regularity in inference, and so on. These are areas of active current research.

While much of the personalized medicine literature is occupied by the use of patients' genetic information in personalizing treatments, the use of genetic information in the dynamic regime context, as of now, is surprisingly limited. Thus we envision this as a critically important research direction in the near future. This being one of the most natural next steps, methodologists will have to carefully investigate how best to handle the associated high dimensionality.

In today's health care, there seems to be an increasing trend in the use of sophisticated mobile devices (e.g. smart phones, actigraph units containing accelerometers, etc.) to remotely monitor patients' chronic health conditions and to act on the fly, when needed. According to the reinforcement learning literature, this is an instance of *online* decision making in a possibly *infinite horizon* setting involving many stages of intervention. Development of statistically sound estimation and inference techniques for such a setting seems to be another very important future research direction.

The call to personalize medicine is growing more urgent, and reaching beyond the walls of academia. Even in popular literature (see, e.g. Topol 2012), it has been declared that

> This is a new era of medicine, in which each person can be near fully defined at the individual level, instead of how we practice medicine at the population level, with [. . .] use of the same medication and dosage for a diagnosis rather than for a patient.

While it is true that high dimensional data, even genome scans, are increasingly available to the average "consumer" of medicine, there remains the need to adequately and appropriately evaluate any new, tailored approach to treatment. It is that evaluation, by statistical means, that has proven theoretically, computationally, and practically challenging and has driven many of the methodological innovations described in this text.

The study of estimation and inference for dynamic treatment regimes is still relatively young, and constantly evolving. Many inferential problems, including inference about the optimal value function, remain incompletely addressed. A further key challenge is the dissemination of the statistical results into the medical and public health spheres, so that the methods being developed are not used in 'toy' examples, but are deployed in routine use for the evidence-based improvement of treatment of chronic illnesses. While observational data can help drive hypotheses and suggest good regimes to explore, increasing the use of SMARTs in clinical research will be required to better understand and evaluate the sequential treatment decisions that are routinely taken in the care of chronic illnesses.