The first few sets of articles (Ref 1.0-1.1) provide some foundational information on semiparametric regression. Semiparametric regression is essentially a combination of parametric and nonparametric models that are used in cases where a nonparametric model may not perform well enough or where a parametric model error distribution is unknown. The three primary semiparametric methods discussed in (Ref 1.0) are the partially linear, index, and varying coefficient models.

Partially linear models are modeled with the following equation: **[INSERT EQUATION]** As can be seen in the equation, there are three separate terms that are used in predicting the dependent variable. The data is assumed to have i.i.d. error with a mean of 0. The model is parametric in the vector B which is simply a vector of all unknown parameters. It is nonparametric in the second term of the equation, where g(Zi) is an unknown function. The goal of this kind of model is to estimate the parameter and nonparametric function in addition to confidence intervals. Robinson’s estimator is an extension of the partial linear model derived through a generalization of residual regression that allows us to concentrate on the unknown g(.) function. This requires that Robinson’s two-step estimator is asymptotically equivalent to the infeasible estimator.

(Ref 1.1) goes on further to discuss the nonparametric and parametric components and their estimation. The author describes feasible nonparametric GLS where the nonparametric nuisance parameter was the variance function. We can then find an estimate for the parameter **[theta]**. The generated regressors model is **[equation]**, where we can use another equation to provide a consistent estimate of the unknown nonparametric parameter through least-squares. The author also describes Andrews MINPIN Theorem, which states that the semiparametric estimator has the same asymptotic distribution when using the true nonparametric component, which holds only when certain assumptions are met. First, the estimators of the parametric and nonparametric components must both be consistent. Other general regularity conditions must also be met (e.g. stochastic equicontinuity, orthogonality between estimates **[theta] and [theta]** In semiparametric estimation, the estimator must typically converge uniformly at some rate, but there is a problem in that some residuals may have unexpected influence on the estimate of our semiparametric parameters. Trimming is one way to approach this issue.

An index model takes the form of **[INSERT EQUATION]** . The data is assumed to have i.i.d. error with a mean of 0. The index model is parametric in its x’B ( a scalar index), and nonparametric in the unknown function g(.) There are two different methods to keep note of regarding the index model. The first, Ichimura’s method, is an index model where the dependent variable y is continuous, and nonlinear least squares is used to estimate Bo and the function g(.) can be estimated using a nonparametric kernel estimator. Klein and Spady’s estimator is an extension of maximum likelihood methods in estimating the parameters. Smoothing coefficient models are another type of semiparametric regression where we have unspecified smooth functions.

(References 2.0-2.3) primarily discusses survival analysis through survival functions, hazard function, PH regression, and censoring. Survival analysis is done to try to model data based on time-to-event, which means censoring will often be a regular part of the data. The survivor function can be interpreted as the equation for the probability that an event of interest has not occurred by a certain time, denoted by **[INSERT SURVIVAL FUNCTION]**. The hazard function is an extension of the survival function in that it denotes the ratio of the probability density function over the survival function, which makes it a conditional density for a given event that has not occurred by a certain time [**INSERT HAZARD FUNCTION]**. Reasons for using hazard functions rather than probability distribution functions are interpretability, simplicity, and modeling ease. As the hazard function’s relationship to time increments are dependent on the given circumstances of the study at hand, the hazard function has a sort of “memoryless” property given separate increments of time that gives the exponential distribution a significant role in hazard regression. As such, some of the families of survival distribution include the exponential, gamma, weibull, and log-normal distributions. Censoring can take several forms: Type 1, Type 2, and random censoring. Type 1 censoring is the situation when censoring time is usually set as a constant while number of survivors in random. Type 2 censoring is more of opposite case, in which time is meant to be used as the random variable until a certain number of survivors is left.

PH regression is based off of the Cox model: **[INSERT COX MODEL]**. In this model, we have two components to keep track of: the baseline hazard function (semiparametric, though must be estimated nonparametrically), and the exponential of the sum of the B parameters and X variable. As with other models, there are some assumptions: The baseline hazard function should not depend on X, only time t; the exponential component involves the random variable X, but not time t; The X variables are time-independent. Cox models are usually fairly robust and fairly simple to interpret because estimated hazards are always non-negative. Cox models usually allow us to define a hazard ratio, which is the ratio of hazard for one individual divided by the hazard for a different individual (allowing us to numerically interpret differences between different levels in survival analysis).

Full likelihood allows us to estimate the baseline hazard function, but the partial likelihood allows us to make estimates of parameter B while accounting for censored data. Once partial likelihood is maximized with respect to the regression parameter, it can then be used to eventually estimate the baseline hazard function. A similar partial likelihood process (penalized partial likelihood) can be utilized in addition to iterative sure independence screening to allow high-dimensional variable selection, which apparently has very small false selection while maintaining small mean squared error. This type of selection through penalization is an extension of classical selection techniques such a stepwise and bootstrap procedures to be used for PH regression, which means simulations can be used to demonstrate the viability of said selection methods. Along with this method of variable selection, Bayesian variable selection can also be performed (though it cannot be readily applied to higher-dimensional selection involving a large number of covariates), in which one would use a nonparametric prior for estimating the baseline hazard function, and a parametric prior is specified for the regression coefficient. Usually, specifying a prior for every parameter in every model is quite complicated, but Bayesian variable selection seems most4 suitable for clinical trials research in which treatments are often modifications of treatment from experiments past.

Along with the models discussed prior, there are other types of flexible models that can model non proportional hazard regression. Regarding variable selection, there are advantages and disadvantages to certain selection methods, including two-step stepwise selection, penalized likelihood, and boosting algorithms (which seems to provide the most practically interpretable effects and ease of reproduction in continued studies).

As a result of modern technology and large volumes of data, machine learning algorithms have been developed, including but not limited to survival trees, neural networks, bayesian methods (as discussed prior), support vector machines, boosting, and other such advanced machine learning techniques. All of these are viable methods for selecting certain predictor variables over others in survival analysis.