Up to now we have been considering models where $X_1, \ldots, X_n$ are i.i.d. with density $p(x; \theta_0)$ in

$$\mathcal{P} = \{p(x; \theta) : \theta \in \Theta \subset R^p\}$$

where $p$ is finite. That is, we have focused on fully parametric models.

In many problems, interest focuses on making inference about a subset of the parameters; however, the entire parameter vector is needed to describe the distribution of the data.
Suppose we are interested in the mean response of some variable, which in a population is normally distributed. So, we conduct an experiment in which we take a random sample of size $n$ from the population. The result is considered a realization of $X_1, \ldots, X_n$, where the $X_i$’s are i.i.d. $N(\mu_0, \sigma_0^2) \in \mathcal{P}$, where

$$\mathcal{P} = \{N(\mu; \sigma^2) : \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+\}$$

Here, $\theta = (\mu, \sigma^2)$. We are interested in estimating $\mu_0$, but $\sigma^2$ is necessary to describe the probability distribution of our data.
In general, it is useful to write the parameter vector $\theta$ as $(\gamma', \psi')'$, where $\gamma$ is the $q$-dimensional parameter vector of interest and $\psi$ is an $r$-dimensional vector of nuisance parameters. In the previous example, $\gamma = \mu$ and $\psi = \sigma^2$. The entire parameter space $\Theta = \Gamma \times \Psi$ has dimension $p = q + r$.

In semiparametric models, the parameter of interest $\gamma$ is still finite-dimensional, but the nuisance parameter $\psi$ is infinite-dimensional, i.e., it cannot be characterized by a finite-dimensional parameter vector. By allowing the nuisance parameter space to be infinite-dimensional, we are placing less restrictions on the probability model from which the data were generated as compared to a fully parametric probability model. Therefore, estimators of $\gamma_0$ in semiparametric models may have greater applicability and greater robustness.
Example 1: Restricted Moment Model

Let $Y$ be a $d$-dimensional random outcome vector and $Z$ a $c$-dimensional covariate vector. Consider the model

$$Y = \mu(Z; \gamma_0) + \epsilon$$

with $E[\epsilon|Z] = 0$. Here, $\mu(Z; \gamma)$ is a specified $d$-dimensional function of $Z$ and $\gamma$, and $\gamma$ is assumed to be $q$-dimensional. The data are $n$ i.i.d. realizations of $X = (Y, Z)$

$$\{X_i = (Y_i, Z_i) : i = 1, \ldots n\}$$
Example 1: Restricted Moment Model

Let $\epsilon(y, z; \gamma) = y - \mu(z; \gamma)$. $(Y, Z) \sim p_{\epsilon, Z}(\epsilon(y, z; \gamma_0), z) \in \mathcal{P}$, where

$$\mathcal{P} = \{p_{\epsilon, Z}(\epsilon(y, z; \gamma), z) = p_{\epsilon|Z}(\epsilon(y, z; \gamma)|z)p_Z(z) : \\
\gamma \in \mathbb{R}^q, p_{\epsilon|Z} \in \Psi_{\epsilon|Z}, p_Z \in \Psi_Z\}$$

Here, the nuisance "parameter" $\psi = (p_{\epsilon|Z}, p_Z)$, where $\Psi_{\epsilon|Z}$ is the space of all conditional densities for $\epsilon$ given $Z$ which have mean 0 and $\Psi_Z$ is the space of all marginal densities for $Z$. 
Take $d = 1$. Assume that the distribution of $\epsilon$ given $Z$ is normally distributed with mean zero and finite variance. Then, $p_{\epsilon,Z}(\epsilon(y, z; \gamma_0), z) \in \mathcal{P}_1$, where

$$\mathcal{P}_1 = \{ p_{\epsilon,Z}(\epsilon(y, z; \gamma), z) = \frac{1}{\sigma} \phi(\epsilon(y, z; \gamma)/\sigma) p_Z(z) : \gamma \in \mathbb{R}^q, \sigma^2 \in \mathbb{R}^+, p_Z \in \Psi_Z \}$$

Here, the nuisance “parameter” $\psi = (\sigma^2, p_Z)$, where $\Psi_Z$ is the space of all marginal densities for $Z$. Note that $\mathcal{P}_1 \subset \mathcal{P}$. 

Example 1: Restricted Moment Model
Example 2: Proportional Hazards Model

Let $T$ and $W$ denote the latent failure and censoring times, respectively. Let $Z$ denote a $k$-dimensional vector of covariates. Under right censoring, the observed data is $X = (O, \Delta, Z)$, where $O = \min(T, W)$ and $\Delta = I(T \leq W)$. We assume that we have $n$ i.i.d. copies of the observed data and that censoring is independent (i.e., $W$ is independent of $T$ given $Z$). The proportional hazards model (Cox, 1972) assumes that

$$
\lambda_{T|Z}(t|Z) = \lambda_0(t) \exp(\gamma_0'Z)
$$

where $\gamma_0$ is the unknown, finite-dimensional parameter of interest and $\lambda_0(t)$ is a unspecified, non-negative function of $t$. 
Example 2: Proportional Hazards Model

Formally, we say that $X \sim p_X(x) \in \mathcal{P}$, where

$$
\mathcal{P} = \left\{ p_X(x; \gamma, \psi) = \lambda_{T|Z}(o|z)^\delta \exp\left(-\int_0^o \lambda_{T|Z}(u|Z)du\right)
\quad \lambda_{C|Z}(o|z)^{1-\delta} \exp\left(-\int_0^o \lambda_{C|Z}(u|Z)du\right)p_Z(z) :
\quad \lambda_{T|Z}(t|z) = \lambda(t) \exp(\gamma'z), \gamma \in \mathbb{R}^q, \lambda \in \Psi_\lambda, \lambda_{C|Z} \in \Psi_{C|Z}, p_Z \in \Psi_Z \right\}
$$

Here, the nuisance “parameter” $\psi = (\lambda, \lambda_{C|Z}, p_Z)$, where $\Psi_\lambda$ is the space of all non-negative function of $t$, $\Psi_{C|Z}$ is the space of all conditional hazard functions for $C$ given $Z$, $\Psi_Z$ is the space of all marginal densities for $Z$. 
Example 2: Proportional Hazards Model

Often researchers impose the additional restriction that $\lambda(t) = \lambda_1 t^{\lambda_2}$ for all $t$ (i.e., Weibull survival regression). Then, $p_X(x) \in \mathcal{P}_1$, where

$$
\mathcal{P}_1 = \{ p_X(x; \gamma, \psi) = \lambda_{T|Z}(o|z)^{\delta} \exp(-\int_0^o \lambda_{T|Z}(u|Z)du) \\
\lambda_{C|Z}(o|z)^{1-\delta} \exp(-\int_0^o \lambda_{C|Z}(u|Z)du)p_Z(z) : \\
\lambda_{T|Z}(t|z) = \lambda_1 t^{\lambda_2} \exp(\gamma'z), \gamma \in \mathbb{R}^q, \lambda_1, \lambda_2 \in \mathbb{R}^+, \lambda_{C|Z} \in \psi_{C|Z}, p_Z \in \psi_Z \}
$$

Here, the nuisance “parameter” $\psi = (\lambda_1, \lambda_2, \lambda_{C|Z}, p_Z)$, where $\psi_{C|Z}$ is the space of all conditional hazard functions for $C$ given $Z$, $\psi_Z$ is the space of all marginal densities for $Z$. Note that $\mathcal{P}_1 \subset \mathcal{P}$. 
A semiparametric estimator, $\hat{\gamma}_n$, of $\gamma_0$ is one which “loosely speaking” has the property that

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{D(\theta_0)} N(0, \Sigma(\theta_0))$$

for all $p(x; \theta_0) \in \mathcal{P}$.

- GEE estimators in restricted moment model
- Maximum partial likelihood estimator in proportional hazards model
Main Questions

1. How do we find semiparametric estimators?
2. How do we find the best semiparametric estimator?
Suppose $X_1, \ldots, X_n$ are i.i.d. with density $p(x; \gamma_0, \psi_0)$ in

$$\mathcal{P} = \{p(x; \gamma, \psi) : \gamma \in \Gamma \subset \mathbb{R}^q, \psi \in \Psi\}$$

where $\Psi$ is an infinite-dimensional space. The parameter $\gamma$ is of primary interest and $\psi$ is a nuisance parameter.
The first step in answering these questions is to consider a finite-parameter model within the semiparametric model. That is, we define a \textit{parametric submodel}.

A parametric submodel is a class of densities

$$\mathcal{P}_s = \{ p(x; \gamma, \nu) : \gamma \in \Gamma \subset \mathbb{R}^q, \nu \in \Upsilon \subset \mathbb{R}^r \}$$

such that

- $\mathcal{P}_s \subset \mathcal{P}$
- $p_0(x) = p(x; \gamma_0, \nu_0) \subset \mathcal{P}_s$. That is, there exists a density in $\mathcal{P}_s$ indexed by $(\gamma_0, \nu_0)$ such that $p_0(x) = p(x; \gamma_0, \nu_0)$. 
Many people get confused between the notion of a *parametric submodel* and a *parametric subclass*. A parametric subclass is a parametric model that is contained within a semiparametric model that we believe will suffice in identifying the probability distribution that generated the data. For example, we may be willing to assume that the conditional distribution of the error given covariates in the restricted model are normally distributed with mean zero and unknown variance $\sigma^2$. This model is contained within the semiparametric model. However, it is not a parametric submodel, if, in truth, the data are not normally distributed.
A parametric submodel is a conceptual idea that is used to help develop a theory for semiparametric models. The reason we say it is conceptual is that we require that a parametric submodel contains the truth. However, since the truth is unknown, we can only describe submodels generically and such submodels are not useful for data analysis.
In the proportional hazards model, we assume

$$
\lambda_{T|Z}(t|Z) = \lambda_0(t) \exp(\gamma_0' Z)
$$

Here is an example of a parametric submodel. Let $h_1(t), \ldots, h_r(t)$ be $r$ different specified functions of $t$. Consider the submodel

$$
\mathcal{P}_s = \{ p_X(x; \gamma, \psi) = \lambda_{T|Z}(o|z)^\delta \exp(-\int_0^o \lambda_{T|Z}(u|Z) du) \\
\lambda_{C|Z}(o|z)^{1-\delta} \exp(-\int_0^o \lambda_{C|Z}(u|Z) du)p_Z(z) : \\
\lambda_{T|Z}(t|z) = \lambda_0(t) \exp(\sum_{j=1}^r \nu_j h_j(t) + \gamma' z), \gamma \in \mathbb{R}^q, (\nu_1, \ldots, \nu_j) \subset \gamma \subset \mathbb{R}^r, \lambda_{C|Z} = \lambda_{C|Z,0}, p_Z = p_{Z,0} \}
$$
Example

- $\mathcal{P}_s$ is a parametric model as it is characterized by $(q + r) < \infty$ parameters.
- $\mathcal{P}_s \subset \mathcal{P}$.
- $p_X(x; \gamma_0, \psi_0) \in \mathcal{P}_s$. Take $\gamma = \gamma_0$ and $\psi = 0$.
- The parametric submodel depends on the unknown $\lambda_0(t)$, unknown hazard of $C$ given $Z$, and unknown density of $Z$, so that such a model is not useful for data analysis.
- Contrast this with $\mathcal{P}_1$ (i.e, Weibull regression model). If we were willing to assume that the true baseline hazard has a parametric form, then we can estimate $\lambda_1$, $\lambda_2$ and $\gamma$ and use this for inference. The downside is that if the form of the baseline hazard is misspecified, then our inference could be misleading.
Previously, we studied RAL estimators of \( \gamma_0 \) in finite-dimensional parametric models and were able to derive asymptotic properties through their influence function. From these results, we know that RAL estimators of \( \gamma_0 \) in a parametric submodel have influence functions that satisfy the following properties:

- The influence functions belong to the orthogonal complement of the nuisance tangent space for the submodel, which is defined as

\[
\Lambda_s = \{ BS_{v,s}(X; \gamma_0, \nu_0) : B^{q \times r} \}
\]

where \( S_{v,s}(X; \gamma_0, \nu_0) = \frac{\partial \log p(X; \gamma_0, \nu_0)}{\partial \nu} \).
Inference in Parametric Submodels

- The most efficient influence function is

\[ \varphi_s^{\text{eff}}(X) = E_{\gamma_0,\psi_0}[S_{\gamma,s}(X; \gamma_0, \nu_0)S_{\gamma,s}(X; \gamma_0, \nu_0)']^{-1}S_{\gamma,s}(X; \gamma_0, \nu_0) \]

where

\[ S_{\gamma,s}(X; \gamma_0, \nu_0) = S_\gamma(X; \gamma_0, \psi_0) - \Pi[S_\gamma(X; \gamma_0, \psi_0)|\Lambda_s] \]

and \( S_{\gamma,s}(X; \gamma_0, \nu_0) = \frac{\partial \log p(X; \gamma_0, \nu_0)}{\partial \gamma} \).

- The smallest variance among such estimators for \( \gamma_0 \) in the parametric submodel is

\[ E_{\gamma_0,\psi_0}[S_{\gamma,s}(X; \gamma_0, \nu_0)S_{\gamma,s}(X; \gamma_0, \nu_0)']^{-1} \]
An estimator of $\gamma_0$ is a RAL estimator for a semiparametric model if it is RAL for every parametric submodel! Therefore, any influence function for a RAL estimator in a semiparametric model must be an influence function for an RAL estimator within a parametric submodel. That is, the class of influence functions for RAL estimators in $\mathcal{P}$ must be a subset of the class of influence functions for RAL estimators in $\mathcal{P}_s$.

A heuristic way of understanding this is as follows. We know that if $\hat{\gamma}_n$ is a semiparametric estimator, then

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{D} N(0, \Sigma(\gamma_0, \psi_0))$$

for all $p(x; \gamma_0, \psi_0) \in \mathcal{P}$. That implies that

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{D} N(0, \Sigma(\gamma_0, \nu_0))$$

for all $p(x; \gamma_0, \nu_0) \in \mathcal{P}_s$. However, the converse may not be true.
Consequently, the class of semiparametric estimators must be contained within the class of estimators for the parametric submodel. Therefore,

1. Any influence function of an RAL semiparametric estimator of \( \gamma_0 \) must be orthogonal to the nuisance tangent spaces from all parametric submodels.

2. The variance of any RAL semiparametric estimator must be greater than or equal to

\[
E_{\gamma_0, \psi_0} [S_{\gamma, s}(X; \gamma_0, \nu_0)S_{\gamma, s}(X; \gamma_0, \nu_0)']^{-1}
\]

for all parametric submodels \( P_s \).

Hence the variance of any semiparametric estimators must be greater than or equal to

\[
\sup_{P_s} E_{\gamma_0, \psi_0} [S_{\gamma, s}(X; \gamma_0, \nu_0)S_{\gamma, s}(X; \gamma_0, \nu_0)']^{-1}
\]
This *supremum* is the *semiparametric efficiency bound*. Any semiparametric estimator, $\hat{\gamma}_n$, with asymptotic variance achieving this bound for a particular $p(x; \gamma_0, \psi_0)$ is said to be *locally efficient*. If the same estimator $\hat{\gamma}_n$ is semiparametric efficient for all $p(x; \gamma_0, \psi_0)$, then we say that such an estimator is *globally semiparametric efficient*. 
Geometrically, the parametric submodel efficient score is the residual of $S_\gamma(X; \gamma_0, \psi_0)$ after projecting it onto to the parametric submodel nuisance tangent space $\Lambda_s$. For a 1-dimensional parameter $\gamma$, the inverse of the norm-squared of this residual is the smallest variance of all influence functions for RAL estimators for the parametric submodel. The analogy can be extended to $q$-dimensional $\gamma$.

As we increase the complexity of the parametric submodel, or consider the linear space spanned by the different nuisance tangent spaces of submodels, the corresponding nuisance tangent space becomes larger and therefore the norm-squared of the residual becomes smaller. Hence, the variance grows larger.
Let's be more formal. We define the *nuisance tangent space for the semiparametric model* as the mean square closure of the set of all elements in our Hilbert space of the form

\[ \mathcal{H} = \{ B_s S_{v,s}(X; \gamma_0, \nu_0) : \text{for all } P_s, B_s \text{ is a conformable matrix with } q \text{ rows} \} \]

The nuisance tangent space is defined as

\[ \Lambda = \{ h(X; \gamma_0, \psi_0) \in \mathcal{H} : E_{\gamma_0,\psi_0}[h(X; \gamma_0, \psi_0)'h(X; \gamma_0, \psi_0)] \text{ there exist a sequence of submodels } s_j \text{ such that } \| h(X; \gamma_0, \psi_0) - B_{s_j} S_{v,s_j}(X; \gamma_0, \nu_{0,j}) \|^2 \xrightarrow{j \to \infty} 0 \} \]
• The efficient score for the semiparametric model is defined as

\[
S_{\gamma}^{\text{eff}}(X; \gamma_0, \psi_0) = S_{\gamma}(X; \gamma_0, \psi_0) - \Pi[S_{\gamma}(X; \gamma_0, \psi_0)|\Lambda]
\]

• The efficient influence function is

\[
\varphi^{\text{eff}}(X) = E_{\gamma_0, \psi_0}[S_{\gamma}^{\text{eff}}(X; \gamma_0, \psi_0)S_{\gamma}^{\text{eff}}(X; \gamma_0, \psi_0)']^{-1}S_{\gamma}^{\text{eff}}(X; \gamma_0, \psi_0)
\]

• Any estimator which is semiparametric and RAL with influence function \( \varphi(X) \) must satisfy

1. \( E_{\gamma_0, \psi_0}[\varphi(X)S_{\gamma}^{\text{eff}}(X; \gamma_0, \psi_0)'] = I \)
2. \( \Pi[\varphi(X)|\Lambda] = 0 \) (i.e., \( \varphi(X) \in \Lambda^\perp \))

• By deriving the orthogonal complement of the nuisance tangent space, we can expect to construct semiparametric, RAL estimators.
Let $Y$ be a $d$-dimensional random outcome vector and $Z$ a $c$-dimensional covariate vector. Consider the model

$$Y = \mu(Z; \gamma_0) + \epsilon$$

with $E[\epsilon|Z] = 0$. Here, $\mu(Z; \gamma)$ is a specified $d$-dimensional function of $Z$ and $\gamma$, and $\gamma$ is assumed to be $q$-dimensional.

The data are $n$ i.i.d. realizations of $X = (Y, Z)$

$$\{X_i = (Y_i, Z_i) : i = 1, \ldots n\}$$
Let \( \epsilon(y, z; \gamma) = y - \mu(z; \gamma) \). \((Y, Z) \sim p_{\epsilon, Z}(\epsilon(y, z; \gamma_0), z) \in \mathcal{P}, \)

where

\[
\mathcal{P} = \{ p_{\epsilon, Z}(\epsilon(y, z; \gamma), z) = p_{\epsilon|Z}(\epsilon(y, z; \gamma)|z)p_Z(z) : \\
\gamma \in R^q, p_{\epsilon|Z} \in \Psi_{\epsilon|Z}, p_Z \in \Psi_Z \}
\]

Here, the nuisance “parameter” \( \psi = (p_{\epsilon|Z}, p_Z) \), where \( \Psi_{\epsilon|Z} \) is the space of all conditional densities for \( \epsilon \) given \( Z \) which have mean 0 and \( \Psi_Z \) is the space of all marginal densities for \( Z \).
A parametric submodel corresponds to parameterizing $p_{\epsilon|Z}$ and $p_Z$ via finite dimensional parameters. So, a parametric submodel is of the form:

$$\mathcal{P}_s = \{p_{\epsilon,Z}(\epsilon(y,z;\gamma),z;\nu) = p_{\epsilon|Z}(\epsilon(y,z;\gamma)|z;\nu_1)p_Z(z;\nu_2) : \\
\gamma \in \mathbb{R}^q, \nu_1 \in \Upsilon_1 \subset \mathbb{R}^{r_1}, \nu_2 \in \Upsilon_2 \subset \mathbb{R}^{r_2}\}$$

where there exists $\nu_0$ such that $p_{\epsilon,Z}(\epsilon,z;\nu_{1,0},\nu_{2,0}) = p_{\epsilon,Z}(\epsilon,z;\psi_0)$
What the log-likelihood for an individual based on the parametric submodel?

\[ l(\gamma, \nu; Y, Z) = \log p_{\epsilon|Z}(Y - \mu(Z; \gamma)|Z; \nu_1) + \log p_Z(Z; \nu_2) \]

The nuisance score for \( \nu \) evaluated at the truth is given by

\[ S_{\nu,s}(X; \gamma_0, \psi_0) = (S_{\nu_1,s}(X; \gamma_0, \psi_0)', (S_{\nu_2,s}(X; \gamma_0, \psi_0)')')' \]

where

\[ S_{\nu_1,s}(X; \gamma_0, \psi_0) = \frac{\partial \log p_{\epsilon|Z}(Y - \mu(Z; \gamma_0)|Z; \psi_0)}{\partial \nu_1} \]
\[ S_{\nu_2,s}(X; \gamma_0, \psi_0) = \frac{\partial \log p_Z(Z; \psi_0)}{\partial \nu_2} \]

Note that \( E_{\gamma_0,\psi_0}[S_{\nu_1,s}(X; \gamma_0, \psi_0)S_{\nu_2,s}(X; \gamma_0, \psi_0)'] = 0. \)
The nuisance tangent space for the submodel can be written as

$$\Lambda_s = \{ B_1 S_{v_1,s}(X; \gamma_0, \psi_0) + B_2 S_{v_2,s}(X; \gamma_0, \psi_0) \} = \Lambda_{1,s} \oplus \Lambda_{2,s}$$

where

$$\Lambda_{1,s} = \{ B_1 S_{v_1,s}(X; \gamma_0, \psi_0) : \forall B_1 \}$$
$$\Lambda_{2,s} = \{ B_2 S_{v_2,s}(X; \gamma_0, \psi_0) : \forall B_2 \}$$

The nuisance tangent space for the semiparametric model, $\Lambda$, is the mean square closure over all $\Lambda_s$. Since $v_1$ and $v_2$ are variation independent, $\Lambda$ can be written as the $\Lambda_1 \oplus \Lambda_2$, where $\Lambda_1$ is the mean square closure over all $\Lambda_{1,s}$ and $\Lambda_2$ is the mean square closure of all $\Lambda_{2,s}$.

- Finding these spaces typically begins with an educated guess, which is then verified.
Deriving $\Lambda_2$

Note that for all parametric submodels $S_{v_2,s}(X; \gamma_0, \psi_0)$ has mean zero and is function only of $Z$. Therefore any element of $\Lambda_{2,s}$ is a $q$-dimensional function of $Z$ with mean zero. It is reasonable to conjecture that $\Lambda_2$, the mean square closure of all $\Lambda_{2,s}$ is the linear subspace of all $q$-dimensional mean zero functions of $Z$. For the moment, call this space $\Lambda_2^c$. Formally,

$$\Lambda_2^c = \{ h(Z); E_{\gamma_0,\psi_0}[h(Z)] = 0 \}$$

In order to demonstrate that our conjecture is true, we must demonstrate that

1. For all $s$, any element of $\Lambda_{2,s}$ belongs to $\Lambda_2^c$. This is obvious
2. Any element of $\Lambda_2^c$ can be written as an element of $\Lambda_{2,s}$, for some $s$. 

Consider any element $h(Z)$ from $\Lambda_2^\xi$. Consider the following parametric submodel

$$\mathcal{P}_s = \{ p_{\epsilon,Z}(\epsilon,z;\nu) = p_{\epsilon|Z}(\epsilon|z;\psi_0)p_Z(z;\psi_0)(1 + \nu'_2 h(z)) : \nu_2 \in \Upsilon_2 \subset \mathbb{R}^q \}$$

where $\Upsilon_2$ is the space of $\nu_2$ sufficiently small such that $(1 + \nu'_2 h(z)) \geq 0$ for all $z$. Here $p_Z(z;\nu_2) = p_Z(z;\psi_0)(1 + \nu'_2 h(z))$ and note that it is a proper density and that taking $\nu_2 = 0$ returns the true density density of $Y$ and $Z$.

Furthermore, note that $S_{\nu_2,s}(X;\gamma_0,\psi_0) = h(Z)$. Thus, we have proven our conjecture - $\Lambda_2 = \Lambda_2^\xi$. 
Deriving $\Lambda_1$

Note that for all parametric submodels $S_{\nu_1,s}(X; \gamma_0, \psi_0)$ has

a. $E_{\gamma_0,\psi_0}[S_{\nu_1,s}(X; \gamma_0, \psi_0)|Z] = 0$.

b. $E_{\gamma_0,\psi_0}[S_{\nu_1,s}(X; \gamma_0, \psi_0)e'|Z] = 0$

Therefore any element of $\Lambda_{1,s}$ is a $q$-dimensional function of $\epsilon$ and $Z$ which satisfy a. and b. It is reasonable to conjecture that $\Lambda_1$, the mean square closure of all $\Lambda_{1,s}$ is the linear subspace of all $q$-dimensional functions of $\epsilon$ and $Z$, which satisfy a. and b.. For the moment, call this space $\Lambda_1^c$. Formally,

$$\Lambda_1^c = \{h(\epsilon, Z) : E_{\gamma_0,\psi_0}[h(\epsilon, Z)|Z] = 0, E_{\gamma_0,\psi_0}[h(\epsilon, Z)e'|Z] = 0\}$$

To prove this conjecture, we must verify that

1. For all $s$, any element of $\Lambda_{1,s}$ belongs to $\Lambda_1^c$. This is obvious.
2. Any element of $\Lambda_1^c$ can be written as an element of $\Lambda_{1,s}$, for some $s$. 
Consider any element \( h(\epsilon, Z) \) from \( \Lambda_1^c \). Consider the following parametric submodel

\[
\mathcal{P}_s = \{ p_{\epsilon,Z}(\epsilon, z; \nu) = p_{\epsilon|Z}(\epsilon|z; \psi_0)(1 + \nu_1' h(\epsilon, z))p_Z(z; \psi_0) : \nu_1 \in \Upsilon_1 \subset \mathbb{R}^q \}
\]

where \( \Upsilon_1 \) is the space of \( \nu_1 \) sufficiently small such that \( (1 + \nu_1' h(\epsilon, z) \geq 0 \) for all \( \epsilon, z \). Here

\[
p_{\epsilon|Z}(\epsilon|z; \nu) = p_{\epsilon|Z}(\epsilon|z; \psi_0)(1 + \nu_1' h(\epsilon, z))
\]

and note that it is a proper conditional density, the mean of \( \epsilon \) given \( Z \) is zero, and that taking \( \nu_1 = 0 \) returns the true density density of \( Y \) and \( Z \).

Furthermore, note that \( S_{\nu_1,s}(X; \psi_0) = h(\epsilon, Z) \). Thus, we have proven our conjecture - \( \Lambda_1 = \Lambda_1^c \).
\[ \Lambda = \Lambda_1 \oplus \Lambda_2, \text{ where} \]

\[ \Lambda_1 = \left\{ h(\epsilon, Z) : E_{\gamma_0, \psi_0}[h(\epsilon, Z)|Z] = 0, E_{\gamma_0, \psi_0}[h(\epsilon, Z)\epsilon'|Z] = 0 \right\} \]

\[ \Lambda_2 = \left\{ h(Z) : E_{\gamma_0, \psi_0}[h(Z)] = 0 \right\} \]

In order to identify influence functions, we need to find \( \Lambda^\perp \). Now,

\[ \Lambda^\perp = \left\{ h(\epsilon, Z) \in \mathcal{H} : \Pi[h(\epsilon, Z)|\Lambda] = 0 \right\} \]

\[ = \left\{ h(\epsilon, Z) - \Pi[h(\epsilon, Z)|\Lambda] : h(\epsilon, Z) \in \mathcal{H} \right\} \]

\[ = \left\{ h(\epsilon, Z) - \Pi[h(\epsilon, Z)|\Lambda_1] - \Pi[h(\epsilon, Z)|\Lambda_2] : h(\epsilon, Z) \in \mathcal{H} \right\} \]
\[ \Pi[h(\epsilon, Z)|\Lambda_2] = E_{\gamma_0, \psi_0}[(h(\epsilon, Z)|Z] \]

We must find \( h^*(Z) \) with mean zero, such that
\[ E_{\gamma_0, \psi_0}[h(\epsilon, Z) - h^*(Z)]h(Z)] = 0 \] for all \( h(Z) \in \Lambda_2 \). The answer is
\[ h^*(Z) = E_{\gamma_0, \psi_0}[(h(\epsilon, Z)|Z]. \]

\[ \Pi[h(\epsilon, Z)|\Lambda_1] = \\
\quad h(\epsilon, Z) - E_{\gamma_0, \psi_0}[h(\epsilon, Z)\epsilon'|Z] Var_{\gamma_0, \psi_0}[\epsilon|Z]^{-1}\epsilon - E_{\gamma_0, \psi_0}[h(\epsilon, Z)|Z] \]

Thus,
\[ \Lambda^\perp = \{ E_{\gamma_0, \psi_0}[h(\epsilon, Z)\epsilon'|Z] Var_{\gamma_0, \psi_0}[\epsilon|Z]^{-1}\epsilon : h(\epsilon, Z) \in \mathcal{H} \} \]
**Definition:** A transformation, $f$, from a vector space $\mathcal{H}$ into the space of real numbers is called a *functional* on $\mathcal{H}$.

**Definition:** An operator, $A$, is a mapping from one space into another. If $A$ maps from $\mathcal{H}$ into $\mathcal{G}$, we write $A : \mathcal{H} \rightarrow \mathcal{G}$. The collection of all vectors $g \in \mathcal{G}$ for which there is an $h \in \mathcal{H}$ such that $g = A(h)$ is called the range of $A$, $\mathcal{R}(A)$. The collection of all vectors $h \in \mathcal{H}$ such that $A(h) = 0$ is called the null space of $A$, $\mathcal{N}(A)$. 
We will be concerned with bounded, linear operators.

**Definition:** An operator $A$ is linear if for any $h_1, h_2 \in \mathcal{H}$, $A(h_1 + h_2) = A(h_1) + A(h_2)$.

**Definition:** A linear operator $A$ from a normed space $\mathcal{H}$ to a normed space $\mathcal{G}$ is said to be bounded if there exists a constant $M$ such that $\|A(h)\| \leq M\|h\|$ for all $h \in \mathcal{H}$, The smallest such $M$ which satisfies this condition is denoted the norm of $A$, $\|A\|$.
Definition: Let $\mathcal{H}$ and $\mathcal{G}$ be normed spaces (i.e., Hilbert spaces) and $A$ be a bounded, linear operator, The adjoint operator $A^* : \mathcal{G}^* \to \mathcal{H}^*$ is defined by the equation

$$< h, A^*(g^*) > = < A(h), g^* >$$

where $< x, y >$ is a bounded, linear functional for each fixed $y$ or each fixed $x$. In our setting, $< x, y > = E[y'x]$.

Theorem: Let $\mathcal{H}$ and $\mathcal{G}$ be Hilbert spaces and let $A$ be a bounded, linear operator mapping from $\mathcal{H}$ to $\mathcal{G}$. Suppose that the closure of $\mathcal{R}(A) \subset \mathcal{G}$. Then for a fixed $g \in \mathcal{G}$, the vector $h_{opt} \in \mathcal{H}$ minimizes $\| g - A(h_{opt}) \|$ if and only if $A^*(A(h_{opt})) = A^*(g)$. $A(h_{opt})$ is the projection of $g$ onto the closure of $\mathcal{R}(A)$.
\[ \Lambda_1 = \{ h(\epsilon, Z) : E_{\gamma_0, \psi_0}[h(\epsilon, Z)|Z] = 0, E_{\gamma_0, \psi_0}[h(\epsilon, Z)\epsilon'|Z] = 0 \} \]
\[ = \{ A(h(\epsilon, Z)) : h(\epsilon, Z) \in \mathcal{H} \} \]

where
\[ A(h(\epsilon, Z)) = A_2(A_1(h(\epsilon, Z))) \]

and
\[ A_1(h(\epsilon, Z)) = h(\epsilon, Z) - E_{\gamma_0, \psi_0}[h(\epsilon, Z)|Z] \]
\[ A_2(h(\epsilon, Z)) = h(\epsilon, Z) - E_{\gamma_0, \psi_0}[h(\epsilon, Z)\epsilon'|Z] \text{Var}_{\gamma_0, \psi_0}[\epsilon|Z]^{-1}\epsilon \]
\[ A^* = A_1^* \cdot A_2^* \]

What is \( A_1^* \)?

\[
\begin{align*}
E[g(\epsilon, Z)'(h(\epsilon, Z) - E[h(\epsilon, Z)|Z])] &= E[A_1^*(g(\epsilon, Z))'h(\epsilon, Z)] \\
E[g(\epsilon, Z)'h(\epsilon, Z)] - E[g(\epsilon, Z)'E[h(\epsilon, Z)|Z]] &= E[A_1^*(g(\epsilon, Z))'h(\epsilon, Z)] \\
E[g(\epsilon, Z)'h(\epsilon, Z)] - E[E[g(\epsilon, Z)'|Z]h(\epsilon, Z)] &= E[A_1^*(g(\epsilon, Z))'h(\epsilon, Z)] \\
E[(g(\epsilon, Z) - E[g(\epsilon, Z)|Z])'h(\epsilon, Z)] &= E[A_1^*(g(\epsilon, Z))'h(\epsilon, Z)] \\
A_1^*(g(\epsilon, Z)) &= g(\epsilon, Z) - E[g(\epsilon, Z)|Z]
\]
What is $A_2^*$?

\[
E[g(\epsilon, Z)'(h(\epsilon, Z) - E[h(\epsilon, Z)|Z] \cdot Var[\epsilon|Z]^{-1}\epsilon)] = E[A_1^*(g(\epsilon, Z))' h(\epsilon, Z)]
\]

\[
E[g(\epsilon, Z)' h(\epsilon, Z)] - E[g(\epsilon, Z)' E[h(\epsilon, Z)|Z] \cdot Var[\epsilon|Z]^{-1}\epsilon] = E[A_1^*(g(\epsilon, Z))' h(\epsilon, Z)]
\]

\[
E[g(\epsilon, Z)' h(\epsilon, Z)] - E[\epsilon' Var[\epsilon|Z]^{-1} E[\epsilon h(\epsilon, Z)'|Z] g(\epsilon, Z)] = E[A_1^*(g(\epsilon, Z))' h(\epsilon, Z)]
\]

\[
E[g(\epsilon, Z)' h(\epsilon, Z)] - E[\epsilon' Var[\epsilon|Z]^{-1} E[\epsilon g(\epsilon, Z)'|Z] h(\epsilon, Z)] = E[A_1^*(g(\epsilon, Z))' h(\epsilon, Z)]
\]

\[
E[(g(\epsilon, Z) - E[g(\epsilon, Z)|Z] \cdot Var[\epsilon|Z]^{-1}\epsilon)' h(\epsilon, Z)] = E[A_1^*(g(\epsilon, Z))' h(\epsilon, Z)]
\]

\[
A_1^*(g(\epsilon, Z)) = g(\epsilon, Z) - E[g(\epsilon, Z)|Z] \cdot Var[\epsilon|Z]^{-1}\epsilon
\]
So,

\[ A^*(g(\epsilon, Z)) = g(\epsilon, Z) - E_{\gamma_0, \psi_0}[g(\epsilon, Z)\epsilon'|Z] \text{Var}_{\gamma_0, \psi_0}[\epsilon|Z]^{-1}\epsilon - E_{\gamma_0, \psi_0}[g(\epsilon, Z)|Z] \]

Note that \( A^* = A \) (i.e., \( A \) is self-adjoint).

So, what is \( \Pi[h(\epsilon, Z)|\Lambda_1] \)?

\[
\Pi[h(\epsilon, Z)|\Lambda_1] = h(\epsilon, Z) - E_{\gamma_0, \psi_0}[h(\epsilon, Z)\epsilon'|Z] \text{Var}_{\gamma_0, \psi_0}[\epsilon|Z]^{-1}\epsilon - E_{\gamma_0, \psi_0}[h(\epsilon, Z)|Z]
\]

\[
\Pi[h(\epsilon, Z)|\Lambda_1] = h(\epsilon, Z) - E_{\gamma_0, \psi_0}[h(\epsilon, Z)\epsilon'|Z] \text{Var}_{\gamma_0, \psi_0}[\epsilon|Z]^{-1}\epsilon - E_{\gamma_0, \psi_0}[h(\epsilon, Z)|Z]
\]
\[
\Lambda^\perp = \{ E_{\gamma_0, \psi_0}[h(\epsilon, Z)\epsilon' | Z] \operatorname{Var}_{\gamma_0, \psi_0}[\epsilon | Z]^{-1} \epsilon : h(\epsilon, Z) \in \mathcal{H} \}
\]

We can also write

\[
\Lambda^\perp = \{ A(Z; \gamma_0, \psi_0) \epsilon : \text{for all } A(Z; \gamma_0, \psi_0) \}
\]

This is because, for given \( A(Z; \gamma_0, \psi_0) \) there exists an \( h(\epsilon, Z) \in \mathcal{H} \) such that \( E_{\gamma_0, \psi_0}[h(\epsilon, Z)\epsilon' | Z] \operatorname{Var}_{\gamma_0, \psi_0}[\epsilon | Z]^{-1} = A(Z; \gamma_0, \psi_0) \).

Specifically, take \( h(\epsilon, Z) = A(Z; \gamma_0, \psi_0) \epsilon \).
What is the Efficient Influence Function?

To find the most efficient semiparametric estimator, it suffices to find the efficient influence function. For this, we need to derive the efficient score, which the residual from the projection of $S_\gamma(X; \gamma_0, \psi_0)$ onto $\Lambda$. So,

$$S^{\text{eff}}_\gamma(X; \gamma_0, \psi_0) = E_{\gamma_0, \psi_0}[S_\gamma(X; \gamma_0, \psi_0)\epsilon'|Z] \text{Var}_{\gamma_0, \psi_0}[\epsilon|Z]^{-1}\epsilon$$

What is $S_\gamma(X; \gamma_0, \psi_0)$? To compute, we fix the nuisance parameters and vary $\gamma$ around $\gamma_0$. So,

$$S_\gamma(X; \gamma_0, \psi_0) = -\frac{\partial p_{\epsilon|Z}(\epsilon|Z)/\partial \epsilon}{p_{\epsilon|Z}(\epsilon|Z)}(\partial \mu(Z; \gamma_0)/\partial \gamma)$$
What is the Efficient Influence Function?

What is $E_{\gamma_0, \psi_0}[S_\gamma(X; \gamma_0, \psi_0)\epsilon'|Z]$. Because of the model restrictions, we know that

$$\int (y - \mu(z; \gamma))p_{\epsilon|Z}(y - \mu(z; \gamma)|Z)dy = 0 \text{ for all } z, \gamma$$

$$\frac{\partial}{\partial \gamma} \int (y - \mu(z; \gamma))p_{\epsilon|Z}(y - \mu(z; \gamma)|Z)dy|_{\gamma_0, \psi_0} = 0 \text{ for all } z, \gamma$$

$$\int (-\partial \mu(z; \gamma_0)/\partial \gamma))p_{\epsilon|Z}(\epsilon|Z))dy + \int \epsilon S_\gamma(y, z; \gamma_0, \psi_0)'dy = 0 \text{ for all } z, \gamma$$

This implies that $E_{\gamma_0, \psi_0}[S_\gamma(X; \gamma_0, \psi_0)\epsilon'|Z] = \partial \mu(z; \gamma_0)/\partial \gamma$. Thus,

$$S_{\gamma_0}^{\text{eff}}(X; \gamma_0, \psi_0) = (\partial \mu(z; \gamma_0)/\partial \gamma) \text{Var}_{\gamma_0, \psi_0}[\epsilon|Z]^{-1}\epsilon$$
Aside: Constructing Estimators

Can we show how take elements from \( \Lambda^\perp \) to construct an estimating equation, whose solution is regular and asymptotically linear. The random vectors in \( \Lambda^\perp \) depend on \( \theta \). Consider any particular element in \( \Lambda^\perp \), say \( h(Z; \theta) \). Now, we know that

\[
E_{\theta_0}[h(Z; \gamma_0, \psi_0)] = 0
\]

whatever be \( \gamma_0 \in \Gamma \) and \( \psi_0 \in \Psi \).

For given \( \gamma \), let \( \hat{\psi}_n(\gamma) \) be a profile estimator of \( \psi \). Assume that

- \( \hat{\psi}_n(\gamma) \) is continuously, differentiable function of \( \gamma \) in a neighborhood of \( \gamma_0 \) and \( \frac{\partial \hat{\psi}_n(\gamma_0)}{\partial \gamma} \) converges in probability to \( d(\theta_0) \).
- For some \( \epsilon > 0 \), \( n^{1/4+\epsilon}(\hat{\psi}_n(\gamma_0) - \psi_0) \) is bounded in probability.
For example, suppose that for fixed $\gamma$, $\hat{\psi}_n(\gamma)$ solves $\sum_{i=1}^n m(Z_i, \gamma, \psi) = 0$, where $m(Z; \gamma, \psi)$ is $q$-dimensional estimating function and $E_{\theta_0}[m(Z; \gamma_0, \psi_0)] = 0$. It can be shown that $\sqrt{n}(\hat{\psi}_n(\gamma_0) - \psi_0)$ is bounded in probability. By the implicit function theorem, we know

$$0 = \frac{\partial \sum_{i=1}^n m(Z_i, \gamma_0, \hat{\psi}_n(\gamma_0))}{\partial \gamma} = \sum_{i=1}^n \frac{\partial m(Z_i, \gamma_0, \hat{\psi}_n(\gamma_0))}{\partial \gamma} + \sum_{i=1}^n \frac{\partial m(Z_i, \gamma_0, \hat{\psi}_n(\gamma_0))}{\partial \psi} \frac{\partial \hat{\psi}_n(\gamma_0)}{\partial \gamma}.$$
This implies that

\[
\frac{\hat{\psi}_n(\gamma_0)}{\partial \gamma} = - \left[ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial m(Z_i, \gamma_0, \hat{\psi}_n(\gamma_0))}{\partial \psi} \right]^{-1} \frac{1}{n} \sum_{i=1}^{n} \frac{\partial m(Z_i, \gamma_0, \hat{\psi}_n(\gamma_0))}{\partial \gamma} 
\]

\[ \xrightarrow{P} E_{\theta_0} \left[ \frac{\partial m(Z, \gamma_0, \psi_0)}{\partial \psi} \right]^{-1} E_{\theta_0} \left[ \frac{\partial m(Z, \gamma_0, \psi_0)}{\partial \gamma} \right] \equiv d(\theta_0) \]
Aside: Constructing Estimators

Now, estimate $\gamma_0$ as the solution, $\hat{\gamma}_n$, to

$$\sum_{i=1}^{n} h(Z_i; \gamma, \hat{\psi}_n(\gamma)) = 0$$

What is the influence function for $\hat{\gamma}_n$?

By standard Taylor series expansions, we know that

$$0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(Z_i; \hat{\gamma}_n, \hat{\psi}_n(\hat{\gamma}_n))$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(Z_i; \gamma_0, \hat{\psi}_n(\gamma_0)) +$$

$$\left[ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial h(Z_i; \gamma_0, \hat{\psi}_n(\gamma_0))}{\partial \gamma} + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial h(Z_i; \gamma_0, \hat{\psi}_n(\gamma_0))}{\partial \psi} \frac{\partial \hat{\psi}_n(\gamma_0)}{\partial \gamma} \right] \cdot \sqrt{n}(\hat{\gamma}_n - \gamma_0) + o_P(1)$$
Now, let’s establish a few facts. First, we can show that

$$
\frac{1}{n} \sum_{i=1}^{n} \frac{\partial h(Z_i; \gamma_0, \hat{\psi}_n(\gamma_0))}{\partial \gamma} \xrightarrow{P} E_{\theta_0} \left[ \frac{\partial h(Z_i; \gamma_0, \psi_0)}{\partial \gamma} \right] = -E_{\theta_0} \left[ h(Z; \theta_0) S_{\gamma}(Z; \theta_0)' \right]
$$

Second, we can show that

$$
\frac{1}{n} \sum_{i=1}^{n} \frac{\partial h(Z_i; \gamma_0, \hat{\psi}_n(\gamma_0))}{\partial \psi} \xrightarrow{P} E_{\theta_0} \left[ \frac{\partial h(Z_i; \gamma_0, \psi_0)}{\partial \psi} \right] = -E_{\theta_0} \left[ h(Z; \theta_0) S_{\psi}(Z; \theta_0)' \right] = 0
$$
Aside: Constructing Estimators

Third, we can show that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(Z_i; \gamma_0, \hat{\psi}_n(\gamma_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(Z_i; \gamma_0, \psi_0) +
\]

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{\partial h(Z_i; \gamma_0, \psi_0)}{\partial \psi} \cdot \sqrt{n}(\hat{\psi}_n(\gamma_0) - \gamma_0) +
\]

\[
\frac{1}{2n} \sum_{i=1}^{n} \frac{\partial^2 h(Z_i; \gamma_0, \psi_0)}{\partial \psi^2} \cdot (n^{1/4}(\hat{\psi}_n(\gamma_0) - \gamma_0))^2 + o_P(1)
\]

Since \( n^{1/4}(\hat{\psi}_n(\gamma_0) - \gamma_0) \) converges in probability to zero and
\[
\frac{1}{n} \sum_{i=1}^{n} \frac{\partial h(Z_i; \gamma_0, \psi_0)}{\partial \psi}
\]
converges in probability to zero, we see that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(Z_i; \gamma_0, \hat{\psi}_n(\gamma_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(Z_i; \gamma_0, \psi_0) + o_P(1)
\]
Putting all these results together, we now conclude that

\[ \sqrt{n}(\hat{\gamma}_n - \gamma_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E_{\theta_0}[\frac{\partial h(Z_i; \gamma_0, \psi_0)}{\partial \gamma}]^{-1} h(Z_i; \gamma_0, \psi_0) \]

So, the influence function for \( \hat{\gamma}_n \) is \( E_{\theta_0}[\frac{\partial h(Z; \gamma_0, \psi_0)}{\partial \gamma}]^{-1} h(Z; \gamma_0, \psi_0) \), which is the influence function that we would have derived if we had pretended as if \( \psi_0 \) were known.
Aside: Constructing Estimators

1. Select an element from the orthogonal complement of the nuisance tangent space.
2. Find an (profile) estimator for the nuisance parameter.
3. The influence function for the resulting estimator will be the same as if the nuisance parameter had been known.
4. This procedure carries over to the case where the nuisance parameter is infinite dimensional!
The optimal estimating function is

\[ \left( \frac{\partial \mu(z; \gamma_0)}{\partial \gamma} \right) \text{Var}_{\gamma_0, \psi_0} \left[ (Y - \mu(Z; \gamma_0)) | Z \right]^{-1} (Y - \mu(Z; \gamma_0)) \]

The optimal estimating function depends on \( \text{Var}_{\gamma_0, \psi_0} \left[ (Y - \mu(Z; \gamma_0)) | Z \right] \), which is difficult to estimate when \( Z \) is high-dimensional. In this situation, we posit some lower-dimensional model for \( \text{Var}_{\gamma_0, \psi_0} \left[ (Y - \mu(Z; \gamma_0)) | Z \right] \) whose parameters are estimated (usually via another set of estimating functions). Consequently, solution to the “optimal” estimating function lead to locally efficient estimators of \( \gamma_0 \). That is, if the posited model for the covariance structure is correct, then the resulting estimator is semiparametric efficient; otherwise it is not. However, even if the model is incorrect, the resulting estimator for \( \gamma_0 \) will be CAN.