Section 10: Role of influence functions in characterizing large sample efficiency

1. Recall that large sample efficiency (of the MLE) can refer only to a class of “regular” estimators.

2. To show this here, we will first show a characterization of a class of “regular” estimators through its “influence function”.

3. To show this characterization, we will use the concept of “contiguity”.
Section 10.1 Asymptotically linear estimators

Let \( X_n = (X_1, X_2, \ldots, X_n) \), where the \( X_i \)'s are i.i.d. \( p(x; \theta_0) \in \mathcal{P} = \{ p(x; \theta) : \theta \in \Theta \} \). Suppose that we are interested in estimating \( \gamma(\theta) \), where \( \gamma(\cdot) : \Theta \to \mathbb{R}^k \) and \( \Theta \subset \mathbb{R}^q \). Most estimators that we will consider are asymptotically linear. That is, there exists a random vector \( \varphi(x) \) such that

\[
\sqrt{n}(\hat{\gamma}(X_n) - \gamma_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(X_i) + o_P(1)
\]

with \( E_{\theta_0}[\varphi(X)] = 0 \) and \( E_{\theta_0}[\varphi(X)\varphi(X)'] \) is finite and non-singular.
The function \( \varphi(x) \) is referred to as an influence function. The phrase influence function was used by Hampel (JASA, 1974) and is motivated by the fact that to the first order \( \varphi(x) \) is the influence of a single observation on the estimator \( \hat{\gamma}(X_n) \).

Consider \( \hat{\gamma}(x_1, \ldots, x_{n-1}, x) \) as a function of \( x \). Since
\[
\sqrt{n}(\hat{\gamma}(x_n) - \gamma_0) \approx \frac{1}{\sqrt{n}} \left( \sum_{i=1}^{n-1} \varphi(x_i) + \varphi(x) \right),
\]
we see that
\[
\hat{\gamma}(x_n) \approx \gamma_0 + \frac{1}{n} \sum_{i=1}^{n-1} \varphi(x_i) + \frac{1}{n} \varphi(x)
\]
The asymptotic properties of an asymptotically linear estimator, \( \hat{\gamma}(X_n) \) can be summarized by considering only its influence function.

Since \( \varphi(X) \) has mean zero, the CLT tells us that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(X_i) \xrightarrow{D} N(0, E_{\theta_0}[\varphi(X)\varphi(X)'])
\]

By Slutsky’s theorem, we know that

\[
\sqrt{n}(\hat{\gamma}(X_n) - \gamma_0) \xrightarrow{D} N(0, E_{\theta_0}[\varphi(X)\varphi(X)'])
\]
Influence Function for GMM Estimators

From the previous section, we know that

\[
\sqrt{n} (\hat{\gamma}(X_n) - \gamma_0) = -\{\hat{D}(\hat{\gamma}(X_n); X_n)'\hat{W}_n D_n^*(X_n)\}^{-1} \hat{D}(\hat{\gamma}(X_n); X_n)'\hat{W}_n \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_i; \gamma_0)
\]

\begin{align*}
&= -\{D'WD\}^{-1} D'W \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_i; \gamma_0) + o_P(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} -\{D'WD\}^{-1} D'W g(X_i; \gamma_0) + o_P(1)
\end{align*}

Hence, the influence function is

\[
\varphi(x) = -\{D'WD\}^{-1} D'W g(x; \gamma_0)
\]

When the dimension of \( g \) is the same as \( \gamma \) and \( D \) and \( W \) are full rank, the influence function is

\[
\varphi(x) = -D^{-1} g(x; \gamma_0)
\]
Example: GEE

The GEE is the solution to

$$\sum_{i=1}^{n} A(X_i; \gamma)(Y_i - \mu(X_i; \gamma)) = 0$$

As a special case of a GMM estimator, we showed that

$$\sqrt{n}(\hat{\gamma}(X_n) - \gamma_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} -E_{\theta_0}[A(X; \gamma_0)M(X; \gamma_0)]^{-1}A(X_i; \gamma_0)(Y_i - \mu(X_i; \gamma_0)) + o_P(1)$$

Therefore, the influence function is

$$\varphi(y, x) = -E_{\theta_0}[A(X; \gamma_0)M(X; \gamma_0)]^{-1}A(x; \gamma_0)(y - \mu(x; \gamma_0))$$

The influence function associated with the optimal GEE is

$$\varphi(y, x) = -E_{\theta_0}[M(X; \gamma_0)'Var_{\theta_0}[Y|X]^{-1}M(X; \gamma_0)]^{-1}M(x; \gamma_0)'Var_{\theta_0}[Y|x]^{-1}(y - \mu(x; \gamma_0))$$
Lemma: An asymptotically linear estimator must have a unique (a.s.) influence function.

Proof: Suppose the influence function is not unique. That is, there exists another influence function \( \varphi^*(x) \) such that \( E_{\theta_0}[\varphi^*(X)] = 0 \) and

\[
\sqrt{n}(\hat{\gamma}(X_n) - \gamma_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi^*(X_i) + o_P(1)
\]

Since,

\[
\sqrt{n}(\hat{\gamma}(X_n) - \gamma_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(X_i) + o_P(1)
\]

we know that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{\varphi(X_i) - \varphi^*(X_i)\} = o_P(1)
\]
By the CLT, we know that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{\varphi(X_i) - \varphi^*(X_i)\} \xrightarrow{D} N(0, Var_{\theta_0}[\varphi(X) - \varphi^*(X)])
\]

In order for this limiting distribution to be \(o_P(1)\),

\[
Var_{\theta_0}[\varphi(X) - \varphi^*(X)] = 0 \text{ or } \varphi(X) - \varphi^*(X) = 0 \text{ a.e.}
\]
In parametric models (i.e., $\Theta$ is finite-dimensional), if an estimator is \textit{asymptotically linear} and \textit{regular} (RAL), then we can establish some useful results regarding the geometry of the influence function as well as precisely defining efficiency.

We assume that $\gamma(\theta)$ is continuously differentiable at least in a neighborhood of the truth $\theta_0$. Let $\gamma_0 = \gamma(\theta_0)$. 
Recall the definition of regular estimator (Section 8)

**Definition:** Consider a local data generating process (LDGP), where for each $n$, the data are distributed according to $\theta_n$, where $\sqrt{n}(\theta_n - \theta_0) \to \tau$. An estimator $\hat{\gamma}(X_n)$ is said to be *regular* if for each $\theta_0$, $\sqrt{n}(\hat{\gamma}(X_n) - \gamma(\theta_n))$ has a limiting distribution that does not depend on the LDGP. So, if

$$\sqrt{n}(\hat{\gamma}(X_n) - \gamma_0) \overset{D}{\to} \theta_0 \ N(0, \Sigma)$$

then

$$\sqrt{n}(\hat{\gamma}(X_n) - \gamma(\theta_n)) \overset{D}{\to} \theta_n \ N(0, \Sigma)$$

We say that “convergence is uniform in a shrinking neighborhood of the true parameter value.” Regularity rules out superefficient estimators (see the example due to Hodges).
Section 10.2 Contiguity

Most density functions in the parametric models with the kind of regularity conditions we have imposed have the property that the sequence of distributions $P_n(X_n; \theta_n)$ are contiguous to the sequence of distributions $P_n(X_n; \theta_0)$, where $\sqrt{n}(\theta_n - \theta_0) \to \tau$.

More formally, consider the following definition:

**Definition:** Consider a sequence of probability measures $\{P_n, Q_n\}$ on the sample spaces $\mathcal{X}_n$ with respect to common dominating measures $\mu_n$, defined on the $\sigma$-algebras $\mathcal{F}_n$. Let $\{p_n, q_n\}$ be the associated sequence of densities. If for any sequence of events $\{A_n\}$ with $A_n \in \mathcal{F}_n$,

$$P_n(A_n) \to 0 \Rightarrow Q_n(A_n) \to 0$$

then we say that the densities $\{q_n\}$ are contiguous to the densities $\{p_n\}$.
This is especially useful since contiguity implies that any sequence of random variables converging to zero in $P_n$ probability also converge to zero in $Q_n$ probability. To see this, suppose that $Z_n \xrightarrow{P_n} 0$. This means that for all $\epsilon > 0$,

$$P_n[|Z_n| > \epsilon] \to 0$$

as $n \to \infty$. Contiguity implies that

$$Q_n[|Z_n| > \epsilon] \to 0$$

as $n \to \infty$. Thus $Z_n \xrightarrow{Q_n} 0$. 
A useful sufficient condition for identifying when a sequence of probability densities are contiguous is given by LeCam’s first lemma.

**LeCam’s First Lemma**

First, we give some preliminary notation and remind you about the results of the Neyman-Pearson lemma (see Section 8.3 of Casella and Berger) for finding most powerful tests.

For a fixed sample size $n$, consider $p_n(x_n)$ to be the fixed null hypothesis and $q_n(x_n)$ to be the fixed alternative. The Neyman-Pearson lemma states the following:
For any level $0 \leq \alpha_n \leq 1$, the most powerful test (randomized) is given by the test function $\phi_n$ defined as

$$\phi_n = \begin{cases} 
0 & q_n < k_n p_n \text{ (Do not reject the null)} \\
\xi_n & q_n = k_n p_n \text{ (Reject the null with probability } \xi_n) \\
1 & q_n > k_n p_n \text{ (Reject the null)} 
\end{cases}$$

Such a $\phi_n$ can always be defined so that the level is equal to $\alpha_n$, i.e.,

$$\alpha_n = \int \phi_n dP_n$$

In particular, for any sequence of events $\{A_n\} \in \mathcal{F}_n$, a corresponding sequence of test functions can be constructed so that

$$\alpha_n = P_n(A_n) = \int \phi_n dP_n$$

Clearly, another sequence of level $\alpha_n$ test functions is given by
\[ \phi_n^*(x_n) = I(x_n \in A_n) \text{ since } \alpha_n = P_n(A_n) = \int \phi_n^* dP_n. \]
However, the N-P Lemma tells us that the most powerful test is given by the likelihood ratio test $\phi_n(x_n)$. Therefore,

\[
\text{Power of } \phi_n = \int \phi_n dQ_n \geq \int \phi^*_n dQ_n = \text{Power of } \phi^*_n = Q_n[A_n]
\]

This implies that to show contiguity of $\{q_n\}$ to $\{p_n\}$ it suffices to show that for the test functions $\phi_n$ defined above that

\[
\int \phi_n dP_n = P_n[A_n] \to 0 \Rightarrow \int \phi_n dQ_n \to 0
\]
Next we introduce a sequence of likelihood ratio random variables:

\[ L_n(x_n) = \frac{q_n(x_n)}{p_n(x_n)} \]

More specifically, we assume that

\[ L_n(x_n) = \begin{cases} 
q_n(x_n)/p_n(x_n) & p_x(x_n) > 0 \\
1 & p_n(x_n) = q_n(x_n) \\
\infty & p_n(x_n) = 0; q_n(x_n) > 0 
\end{cases} \]

Let \( F_n \) be the distribution of \( L_n \) under \( P_n \), i.e.,

\[ F_n(u) = P_n[L_n(X_n) \leq u] \]
Lemma 10.1 (LeCam’s First Lemma): Assume that $F_n$ converges to the distribution function $F$ such that $\int_0^\infty u dF(u) = 1$. Then, the densities $\{q_n\}$ are contiguous to the densities $\{p_n\}$.

Proof: Take a sequence of test functions $\phi_n$ such that
\[ \int \phi_n dP_n \to 0. \] Note that

\[
\int \phi_n dQ_n = \int_{L_n \leq y} \phi_n dQ_n + \int_{L_n > y} \phi_n dQ_n
\]

\[
= \int_{L_n \leq y} \phi_n \frac{q_n}{p_n} p_n d\mu + \int_{L_n > y} \phi_n dQ_n
\]

\[
\leq y \int \phi_n dP_n + \int_{L_n > y} dQ_n
\]

\[
= y \int \phi_n dP_n + 1 - \int_{L_n \leq y} dQ_n
\]

\[
= y \int \phi_n dP_n + 1 - \int_{L_n \leq y} L_n dP_n
\]

\[
= y \int \phi_n dP_n + 1 - \int I(L_n \leq y) L_n dP_n
\]

\[
= y \int \phi_n dP_n + 1 - E_{P_n} [I(L_n \leq y) L_n]
\]

\[
= y \int \phi_n dP_n + 1 - \int_0^y u dF_n(u)
\]
Now, we can find a value $y$ such that $1 - \int_0^y udF(u) < \epsilon/4$. Since $F_n \to F$, we know that $\int_0^y udF_n(u) \to \int_0^y udF(u)$. Therefore, there is some $N_0$ such that for all $n > N_0$, $1 - \int_0^y udF_n(u) < \epsilon/2$ (Use triangle inequality). Furthermore, since $\int \phi_n dP_n \to 0$, we know that there exists $N_1$ so that for all $n > N_1$, $y \int \phi_n dP_n < \epsilon/2$. Therefore, $\int \phi_n dQ_n < \epsilon$ for $n > \max(N_0, N_1)$. Thus, we have shown that $\int \phi_n dQ_n \to 0$. Thus, we have contiguity of the sequence $\{q_n\}$ to the sequence $\{p_n\}$. 
Corollary 10.2: If $\log \{ L_n \} \overset{D(P_n)}{\rightarrow} N(-\sigma^2/2, \sigma^2)$, then the densities \{q_n\} are contiguous to the densities \{p_n\}.

Proof: We see that $L_n$ converges in distribution to a log-normal random variable. Verify that the expectation of a log normal random variable with mean $-\sigma^2/2$ and variance $\sigma^2$ is equal to 1.
We can now apply Corollary 10.2 to our i.i.d. situation. We know that
\[ p_n(x_n) = \prod_{i=1}^{n} p(x_i; \theta_0) \] and
\[ q_n(x_n) = \prod_{i=1}^{n} p(x_i; \theta_n), \] where
\[ \sqrt{n}(\theta_n - \theta_0) \to \tau. \] Now,

\[ \log L_n(X_n) = \log \left\{ \frac{q_n(X_n)}{p_n(X_n)} \right\} = \sum_{i=1}^{n} \{ \log p(X_i; \theta_n) - \log p(X_i; \theta_0) \} \]

By mean value expansion, we see that

\[
\sum_{i=1}^{n} \log p(X_i; \theta_n) = \sum_{i=1}^{n} \log p(X_i; \theta_0) + \sum_{i=1}^{n} \frac{\partial \log p(X_i; \theta_0)}{\partial \theta'} (\theta_n - \theta_0) + \frac{1}{2} (\theta_n - \theta_0)' \sum_{i=1}^{n} \frac{\partial^2 \log p(x_i; \theta^*_n)}{\partial \theta \partial \theta'} (\theta_n - \theta_0)
\]
This implies that

$$\log L_n(X_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i; \theta_0) \cdot \sqrt{n} (\theta_n - \theta_0) +$$

$$\frac{1}{2} \sqrt{n} (\theta_n - \theta_0)' \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log p(x_i; \theta_n^*)}{\partial \theta \partial \theta'} \sqrt{n} (\theta_n - \theta_0)$$

$$\overset{D}{\rightarrow} N\left(-\frac{1}{2} \tau' I(\theta_0) \tau, \tau' I(\theta_0) \tau\right)$$

This satisfies the corollary with $\sigma^2 = \tau' I(\theta_0) \tau$, thus proving that the densities $\{\prod_{i=1}^{n} p(x_i; \theta_n)\}$ are contiguous to $\{\prod_{i=1}^{n} p(X_i; \theta_0)\}$.
Section 10.3 Characterization of “regular” asymptotic linear estimators

**Theorem 10.3:** Suppose that \( \hat{\gamma}(X_n) \) is an asymptotically linear estimator with influence function \( \varphi(X) \) and \( E_\theta[\varphi(X)\varphi(X)'] \) exists and is continuous in the neighborhood of \( \theta_0 \). Then, \( \hat{\gamma}(X_n) \) is regular if and only if

\[
\frac{\partial \gamma(\theta_0)}{\partial \theta'} = E_{\theta_0}[\varphi(X)\psi(X;\theta_0)']
\]

**Proof:** By the definition of an influence function, we know that

\[
\sqrt{n}(\hat{\gamma}(X_n) - \gamma_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(X_i) + o_{P(\theta_0)}(1)
\]
That is,

\[
P_{\theta_0} \left[ \sqrt{n} (\hat{\gamma}(X_n) - \gamma_0) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(X_i) \right] > \epsilon \rightarrow 0
\]

By contiguity, we also know that

\[
P_{\theta_n} \left[ \sqrt{n} (\hat{\gamma}(X_n) - \gamma_0) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(X_i) \right] > \epsilon \rightarrow 0
\]

or \( \sqrt{n} (\hat{\gamma}(X_n) - \gamma_0) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(X_i) \) \( P(\theta_n) \rightarrow 0 \). Adding and subtracting similar terms, we find that

\[
\sqrt{n}(\hat{\gamma}(X_n) - \gamma_n) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\varphi(X_i) - E_{\theta_n}[\varphi(X)]) + \sqrt{n}(\gamma_n - \gamma_0) - \sqrt{n}E_{\theta_n}[\varphi(X)] P(\theta_n) \rightarrow 0
\]
or

\[ \sqrt{n}(\hat{\gamma}(X_n) - \gamma_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\varphi(X_i) - E_{\theta_n}[\varphi(X)]) - \sqrt{n}(\gamma_n - \gamma_0) + \sqrt{n}E_{\theta_n}[\varphi(X)] + o_{P(\theta_n)}(1) \]
We can invoke the CLT for triangular arrays to show that (1) converges in distribution to $N(0, E_{\theta_0}[\varphi(X)\varphi(X)'])$ random vector. For (2), we start by using the mean value theorem to show that

$$
\gamma_n = \gamma(\theta_n) = \gamma_0 + \frac{\partial \gamma(\theta^*_n)}{\partial \theta'}(\theta_n - \theta_0)
$$

This implies that

$$
\sqrt{n}(\gamma_n - \gamma_0) = \frac{\partial \gamma(\theta^*_n)}{\partial \theta'}\sqrt{n}(\theta_n - \theta_0) \to \frac{\partial \gamma(\theta_0)}{\partial \theta'}\tau.
$$

For (3), we know that

$$
\sqrt{n}E_{\theta_n}[\varphi(X)] = \sqrt{n} \int \varphi(x)p(x; \theta_n)d\mu(x)
$$

$$
= \sqrt{n} \int \varphi(x)\{p(x; \theta_0) + \frac{\partial p(x; \theta^*_n)}{\partial \theta'}(\theta_n - \theta_0)\}d\mu(x)
$$

$$
= \sqrt{n}\{ \int \varphi(x)p(x; \theta_0)d\mu(x) + \int \varphi(x) \frac{\partial p(x; \theta^*_n)}{\partial \theta'} p(x; \theta_0)(\theta_n - \theta_0)d\mu(x) \}
$$

$$
= \int \varphi(x) \frac{\partial p(x; \theta^*_n)}{p(x; \theta_0)} p(x; \theta_0)\sqrt{n}(\theta_n - \theta_0)d\mu(x)
$$
Thus,

\[ \sqrt{n} E_{\theta_n} [\varphi(X)] \to \int \varphi(x) \psi(x; \theta_0)' \tau p(x; \theta_0) d\mu(x) = E_{\theta_0} [\varphi(X) \psi(X; \theta_0)'] \tau \]

Putting all of these results together, we find that

\[ \sqrt{n} (\hat{\gamma}(X_n) - \gamma_n) \overset{D(\theta_n)}{\to} N (E_{\theta_0} [\varphi(X) \psi(X; \theta_0)'] \tau - \frac{\partial \gamma(\theta_0)}{\partial \theta'} \tau, E_{\theta_0} [\varphi(X) \varphi(X)']) \]

In order for \( \hat{\gamma}(X_n) \) to be regular, we must have

\[ E_{\theta_0} [\varphi(X) \psi(X; \theta_0)'] - \frac{\partial \gamma(\theta_0)}{\partial \theta'} = 0 \]

which gives the desired result. The other direction is obviously true.
Section 10.4 Geometry and Efficiency of influence functions among regular asymptotically linear estimators

Suppose that we parameterize our problem so that $\theta = (\gamma', \lambda')'$. In this case,

$$\frac{\partial \gamma(\theta_0)}{\partial \theta'} = [ [I] \ [0] ] = [ [E_{\theta_0}[\varphi(X)\psi_{\gamma}(X; \theta_0)']] [E_{\theta_0}[\varphi(X)\psi_{\lambda}(X; \theta_0)']] ]$$

Consider the Hilbert space, $\mathcal{H}$, of $k$-dimensional functions of $X$ with mean zero and finite second moments, equipped with inner product $< h_1, h_2 > = E_{\theta_0}[h_1(X)'h_2(X)]$. The nuisance tangent space is the linear subspace

$$\Lambda = \{ B\psi_{\lambda}(X; \theta_0) : B \text{ is an arbitrary } k \times q \text{ matrix of real numbers} \}$$

The theorem tells us that the influence function, $\varphi(X)$, of a RAL estimator, $\hat{\gamma}(X_n)$, must be orthogonal to every element in $\Lambda$. That is, $\varphi(X) \in \Lambda^\perp = \{ h(X) \in \mathcal{H} : E_{\theta_0}[h(X)'B\psi_{\lambda}(X; \theta_0)] = 0 \}$. In
addition, $E_{\theta_0}[\varphi(X)\psi_{\gamma}(X; \theta_0)] = I$. 
Can we show how to take elements from $\Lambda^\perp$ to construct an estimating equation, whose solution is regular and asymptotically linear? As we will see, the random vectors in $\Lambda^\perp$ depend on $\theta_0$. Consider any particular element in $\Lambda^\perp$, say $h(X; \theta_0)$. Now, we know that

$$E_{\gamma_0, \lambda_0}[h(X; \gamma_0, \lambda_0)] = 0$$

whatever be $\gamma_0 \in \Gamma$ and $\lambda_0 \in \lambda$.

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

- $\hat{\lambda}_n(\gamma)$ is continuously, differentiable function of $\gamma$ in a neighborhood of $\gamma_0$ and $\frac{\partial \hat{\lambda}_n(\gamma_0)}{\partial \gamma}$ converges in probability to $d(\theta_0)$.
- For some $\epsilon > 0$, $n^{1/4+\epsilon}(\hat{\lambda}_n(\gamma_0) - \lambda_0)$ is bounded in probability.
For example, suppose that for fixed $\gamma$, $\hat{\lambda}_n(\gamma)$ solves
\[ \sum_{i=1}^{n} m(X_i, \gamma, \lambda) = 0, \] where $m(X; \gamma, \lambda)$ is $q$-dimensional estimating function and $E_{\theta_0}[m(X; \gamma_0, \lambda_0)] = 0$. It can be shown that
\[ \sqrt{n} (\hat{\lambda}_n(\gamma_0) - \lambda_0) \] is bounded in probability. By the implicit function theorem, we know
\[
0 = \frac{\partial}{\partial \gamma} \sum_{i=1}^{n} m(X_i, \gamma_0, \hat{\lambda}_n(\gamma_0)) \\
= \sum_{i=1}^{n} \frac{\partial m(X_i, \gamma_0, \hat{\lambda}_n(\gamma_0))}{\partial \gamma} + \sum_{i=1}^{n} \frac{\partial m(X_i, \gamma_0, \hat{\lambda}_n(\gamma_0))}{\partial \lambda} \frac{\partial \hat{\lambda}_n(\gamma_0)}{\partial \gamma}
\]

This implies that
\[
\frac{\partial \hat{\lambda}_n(\gamma_0)}{\partial \gamma} = -\left[ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial m(X_i, \gamma_0, \hat{\lambda}_n(\gamma_0))}{\partial \lambda} \right]^{-1} \frac{1}{n} \sum_{i=1}^{n} \frac{\partial m(X_i, \gamma_0, \hat{\lambda}_n(\gamma_0))}{\partial \gamma} \\
\overset{P}{\rightarrow} -E_{\theta_0} \left[ \frac{\partial m(X, \gamma_0, \lambda_0)}{\partial \lambda} \right]^{-1} E_{\theta_0} \left[ \frac{\partial m(X, \gamma_0, \lambda_0)}{\partial \gamma} \right] \equiv d(\theta_0)
\]
Now, estimate $\gamma_0$ as the solution, $\hat{\gamma}_n$, to

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

Under regularity conditions, it can be shown that $\hat{\gamma}_n$ is a consistent estimator of $\gamma_0$.

What is the influence function for $\hat{\gamma}_n$? By mean value expansions, we know that

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

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

$$\left[ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial h(X_i; \gamma^*_n, \hat{\lambda}_n(\gamma^*_n))}{\partial \gamma} \right] + \left[ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial h(X_i; \gamma^*_n, \hat{\lambda}_n(\gamma^*_n))}{\partial \lambda} \frac{\partial \hat{\lambda}_n(\gamma^*_n)}{\partial \gamma} \right] \cdot \sqrt{n}(\hat{\gamma}_n - \gamma_0)$$
\[ \sqrt{n}(\hat{\gamma}_n - \gamma_0) = - \left[ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial h(X_i; \gamma^*_n, \hat{\lambda}_n(\gamma^*_n))}{\partial \gamma} + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial h(X_i; \gamma^*_n, \hat{\lambda}_n(\gamma^*_n))}{\partial \lambda} \frac{\partial \hat{\lambda}_n(\gamma^*_n)}{\partial \gamma} \right]^{-1} \times \]
\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(X_i; \gamma_0, \hat{\lambda}_n(\gamma_0)) \]

Now, for each component of \( h(X, \gamma_0, \hat{\lambda}_n(\gamma_0)) \), we know that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_j(X_i; \gamma_0, \hat{\lambda}_n(\gamma_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_j(X_i; \gamma_0, \lambda_0) + \]
\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial h(X_i; \gamma_0, \lambda_0)}{\partial \lambda} \cdot \frac{1}{n^{1/4+\epsilon}} \cdot n^{1/4+\epsilon} (\hat{\lambda}_n(\gamma_0) - \lambda_0) + \]
\[ \frac{1}{2} n^{1/4} (\hat{\lambda}_n(\gamma_0) - \lambda_0)' \{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 h_j(X_i; \gamma_0, \lambda^*_n)}{\partial \lambda \partial \lambda'} \} n^{1/4} (\hat{\lambda}_n(\gamma_0) + \lambda_0)' \]

Thus,

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_j(X_i; \gamma_0, \hat{\lambda}_n(\gamma_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_j(X_i; \gamma_0, \lambda_0) + o_P(1) \] (5)
Plugging (5) into (4), we get

\[
\sqrt{n}(\hat{\gamma}_n - \gamma_0) = - \left[ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial h(X_i; \gamma_n^*, \hat{\lambda}_n(\gamma^*_n))}{\partial \gamma} + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial h(X_i; \gamma_n^*, \hat{\lambda}_n(\gamma^*_n))}{\partial \lambda} \frac{\partial \hat{\lambda}_n(\gamma^*_n)}{\partial \gamma} \right]^{-1} \times \\
\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(X_i; \gamma_0, \lambda_0) + o_P(1) \right\} \tag{6}
\]

Since, we can show that

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{\partial h(X_i; \gamma_n^*, \hat{\lambda}_n(\gamma^*_n))}{\partial \gamma} \xrightarrow{P} E_{\theta_0} \left[ \frac{\partial h(X_i; \gamma_0, \lambda_0)}{\partial \gamma} \right] = -E_{\theta_0} [h(X; \theta_0) \psi_\gamma(X; \theta_0)']
\]

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{\partial h(X_i; \gamma_n^*, \hat{\lambda}_n(\gamma^*_n))}{\partial \lambda} \xrightarrow{P} E_{\theta_0} \left[ \frac{\partial h(X_i; \gamma_0, \lambda_0)}{\partial \lambda} \right] = -E_{\theta_0} [h(X; \theta_0) \psi_\lambda(X; \theta_0)'] = 0
\]

\[
\frac{\partial \hat{\lambda}_n(\gamma^*_n)}{\partial \gamma} \xrightarrow{P} d(\theta_0)
\]
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(X_i; \gamma_0, \lambda_0) + o_P(1) \overset{D}{\longrightarrow} N(0, E_{\theta_0}[h(X_i; \gamma_0, \lambda_0)h(X_i; \gamma_0, \lambda_0)'])
\]

we can conclude that

\[
\sqrt{n}(\hat{\gamma}_n - \gamma_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} -E_{\theta_0} \left[ \frac{\partial h(X_i; \gamma_0, \lambda_0)}{\partial \gamma} \right]^{-1} h(X_i; \gamma_0, \lambda_0) + o_P(1)
\]
Thus,

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) \overset{D}{\rightarrow} N(0, E_{\theta_0} \left[ \frac{\partial h(X_i; \gamma_0, \lambda_0)}{\partial \gamma} \right]^{-1} E_{\theta_0} \left[ h(X_i; \gamma_0, \lambda_0) h(X_i; \gamma_0, \lambda_0)' \right] E_{\theta_0} \left[ \frac{\partial h(X_i; \gamma_0, \lambda_0)}{\partial \gamma} \right]^{-1'})$$

So, the influence function for $\hat{\gamma}_n$ is $-E_{\theta_0} \left[ \frac{\partial h(X_i; \gamma_0, \lambda_0)}{\partial \gamma} \right]^{-1} h(X; \gamma_0, \lambda_0)$, which is the influence function that we would have derived if we had pretended as if $\lambda_0$ were known.
Algorithm

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!
Efficiency

A geometric interpretation can be useful in helping us define the most efficient estimator among the class of regular, asymptotically linear (RAL) estimators. Since the asymptotic variance of a RAL estimator is the variance of its influence function, we can now look for the most efficient estimator by finding the influence function with the smallest variance.

If we have two influence functions, say $\varphi_1(X)$ and $\varphi_2(X)$, we say that

$$Var_{\theta_0}[\varphi_1(X)] \leq Var_{\theta_0}[\varphi_2(X)]$$

if and only if

$$Var_{\theta_0}[a'\varphi_1(X)] \leq Var_{\theta_0}[a'\varphi_2(X)]$$

for $a \in \mathbb{R}^q$. 
Or equivalently,

\[ a' E_{\theta_0} [\varphi_1(X) \varphi_1(X)']a \leq a' E_{\theta_0} [\varphi_2(X) \varphi_2(X)']a \]

which means that

\[ E_{\theta_0} [\varphi_2(X) \varphi_2(X)'] - E_{\theta_0} [\varphi_1(X) \varphi_1(X)'] \]

is positive semi-definite.
**Efficient Score and Efficient Influence Function**

**Definition:** The efficient score for $\gamma$ is defined as the residual from the projection of $\psi_\gamma(X; \theta_0)$ onto $\Lambda$. That is,

$$\psi^\text{eff}_\gamma(X; \theta_0) = \psi_\gamma(X; \theta_0) - \Pi[\psi_\gamma(X; \theta_0)|\Lambda]$$

where

$$\Pi[\psi_\gamma(X; \theta_0)|\Lambda] = E_{\theta_0}[\psi_\gamma(X; \theta_0)\psi_\lambda(X; \theta_0)']E_{\theta_0}[\psi_\lambda(X; \theta_0)\psi_\lambda(X; \theta_0)']^{-1}\psi_\lambda(X; \theta_0)$$

By construction $\psi^\text{eff}_\gamma(X; \theta_0) \in \Lambda^\perp$. By normalizing it so that that its inner product with $\psi_\gamma(X; \theta_0)$ is one, we then define the efficient influence function

$$\varphi^\text{eff}(X) = E_{\theta_0}[\psi^\text{eff}_\gamma(X; \theta_0)\psi_\gamma(X; \theta_0)']^{-1}\psi^\text{eff}_\gamma(X; \theta_0)$$

$$= E_{\theta_0}[\psi^\text{eff}_\gamma(X; \theta_0)\psi^\text{eff}_\gamma(X; \theta_0)']^{-1}\psi^\text{eff}_\gamma(X; \theta_0)$$

We are now in a position to prove that $\varphi^\text{eff}(X)$ is the most
efficient influence function.
**Claim:** All influence functions can be written as
\[ \varphi^{\text{eff}}(X) + l(X) \]
where \( l(X) \perp \Lambda \) and \( l(X) \perp \{ B\varphi^{\text{eff}}(X) : \text{for all } B \} \).

Since \( \varphi^{\text{eff}}(X) \) is an influence function, it is clear that \( \Lambda \perp \{ B\varphi^{\text{eff}}(X) : \text{for all } B \} \). Now,
\[ l(X) \in \Lambda \perp \cap \{ B\varphi^{\text{eff}}(X) : \text{for all } B \} = [\Lambda \oplus \{ B\varphi^{\text{eff}}(X) : \text{for all } B \}] \perp \]

**Proof:** If \( \varphi(X) \) is an influence function, then \( \varphi(X) \in \Lambda \perp \). In addition \( \varphi^{\text{eff}}(X) \in \Lambda \perp \). Therefore, \( l(X) = \varphi(X) - \varphi^{\text{eff}}(X) \in \Lambda \perp \).

Since \( l(X) \in \Lambda \perp \), we know that \( E_{\theta_0}[l(X)\psi^{\text{eff}}_{\gamma}(X; \theta_0)] = 0 \). Since the space spanned by \( \psi^{\text{eff}}_{\gamma}(X; \theta_0) \) is equal to \( \{ B\varphi^{\text{eff}}(X) : \text{for all } B \} \), we have that \( l(X) \in \{ B\varphi^{\text{eff}}(X) : \text{for all } B \} \perp \).
From this, it is now easy to show that $\varphi^{\text{eff}}(X)$ is the influence function with the smallest variance. Note that

$$E_{\theta_0}[\varphi(X)\varphi(X)'] = E_{\theta_0}[(\varphi^{\text{eff}}(X) + l(X))(\varphi^{\text{eff}}(X) + l(X))']$$

$$= E_{\theta_0}[\varphi^{\text{eff}}(X)\varphi^{\text{eff}}(X)] + E_{\theta_0}[l(X)l(X)']$$

Hence, $E_{\theta_0}[\varphi(X)\varphi(X)'] - E_{\theta_0}[\varphi^{\text{eff}}(X)\varphi^{\text{eff}}(X)]$ is positive semi-definite. This implies that $\varphi^{\text{eff}}(X)$ is the most efficient influence function. The variance of the efficient influence function is $E_{\theta_0}[\psi_{\gamma}^{\text{eff}}(X; \theta_0)\psi_{\gamma}^{\text{eff}}(X; \theta_0)']^{-1}$, the inverse of the variance of the efficient score.

Then we get the well known result that the minimum variance for the most efficient, regular estimator is

$$[I_{\gamma\gamma}(\theta_0) - I_{\gamma\lambda}(\theta_0)I_{\lambda\lambda}(\theta_0)^{-1}I_{\gamma\lambda}(\theta_0)']^{-1}$$