Chapter 7: Influence Functions

Daniel O. Scharfstein

10/16/08
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 $\gamma(\cdot)$, where $\gamma(\cdot) : \Theta \to \mathbb{R}^k$.

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

\[
\sqrt{n}(\hat{\gamma}(X_n) - \gamma_0) \xrightarrow{D} N(0, E_{\theta_0}[\varphi(X) \varphi(X)'])
\]
Lemma 7.1: 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

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{\varphi(X_i) - \varphi^*(X_i)\} = o_P(1)$$
Uniqueness

By the CLT, we know that

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

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

\( \text{Var}_{\theta_0}[\varphi(X) - \varphi^*(X)] = 0 \) or \( \varphi(X) - \varphi^*(X) = 0 \) a.e.
In parametric models (i.e., $\Theta$ is finite-dimensional), if an estimator is *asymptotically linear* and *regular* (RAL), then we can establish some nice 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 $\gamma_0 = \gamma(\theta_0)$. 
In order to avoid problems of *super-efficiency*, we will also insist that our estimator is *regular*.

**Definition 7.1:** 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

$$\sqrt{n}(\hat{\gamma}(X_n) - \gamma_0) \overset{D(\theta_0)}{\to} N(0, \Sigma_0)$$

then

$$\sqrt{n}(\hat{\gamma}(X_n) - \gamma(\theta_n)) \overset{D(\theta_n)}{\to} N(0, \Sigma_0)$$

We say that “convergence is uniform in a shrinking neighborhood of the true parameter value.”
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 7.2:** 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 $Z_n \overset{P_n}{\rightarrow} 0$. This means that for all $\epsilon > 0$,

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

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

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

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

First, we given some preliminary notation and remind you about the results of the Neyman-Pearson lemma 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 $\phi_n^*(x_n) = I(x_n \in A_n)$ since $\alpha_n = P_n(A_n) = \int \phi_n^* dP_n$. 

Neyman-Pearson Lemma
However, the N-P Lemma tells us that the most powerful test is given by the likelihood ratio test \( \phi_n(x_n) \). Therefore,

\[
\text{Power of } \phi_n = \int \phi_n dQ_n \geq \int \phi_n^* dQ_n = \text{Power of } \phi_n^*
\]

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

\[
\int \phi_n dP_n \to 0 \Rightarrow \int \phi_n dQ_n \to 0
\]
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 7.2: 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 $\{q_n\}$.

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

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

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

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

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

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

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

\[
= y \int \phi_n dP_n + 1 - \int_0^y u dF_n(u)
\]
Now, we can find a value $y$ such that $1 - \int_0^y u \, dF(u) < \epsilon/4$. Since $F_n \to F$, we know that $\int_0^y u \, dF_n(u) \to \int_0^y u \, dF(u)$. Therefore, there is some $N_0$ such that for all $n > N_0$, $1 - \int_0^y u \, dF_n(u) < \epsilon/2$ (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 $\{p_n\}$. 

LeCam’s First Lemma
Corollary 7.3: If \( \log\{L_n\} \xrightarrow{D} N(-\sigma^2/2, \sigma^2) \), 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 know apply Corollary 7.3 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,

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

By mean value expansion, we see that

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

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

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

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

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 \( \{ \prod_{i=1}^{n} p(X_i; \theta_0) \} \).
**Theorem 7.4:** 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,

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

By contiguity, we also know that

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

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

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

or

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

(1)
Influence Functions

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 \( \sqrt{n}(\gamma_n - \gamma_0) = \frac{\partial \gamma(\theta_n^*)}{\partial \theta'}(\theta_n - \theta_0) \rightarrow \frac{\partial \gamma(\theta_0)}{\partial \theta'} \cdot \tau \). For (3), we know that

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

\[
= \int \varphi(x)\{p(X; \theta_0) + \frac{\partial p(x; \theta_n^*)}{\partial \theta'}(\theta_n - \theta_0)\}d\mu(x)
\]

\[
= \int \varphi(x)p(X; \theta_0)d\mu(x) + \int \varphi(x)\frac{\partial p(x; \theta_n^*)}{\partial \theta'}p(x; \theta_0)(\theta_n - \theta_0)d\mu(x)
\]

\[
= \int \varphi(x)\frac{\partial p(x; \theta_n^*)}{\partial \theta'}\frac{p(x; \theta_0)}{p_x(x; \theta_0)}p(x; \theta_0)(\theta_n - \theta_0)d\mu(x)
\]
Thus,

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

Putting all of these results together, we find that

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

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.
Geometry of Influence Functions

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

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

If we define the nuisance tangent space to be the linear space spanned by the elements of the nuisance score vector, then every element of the influence function \( \varphi(X) \) is orthogonal to the nuisance tangent space, i.e.,

\[
E_{\theta_0}[\varphi_j(X)\psi_\lambda(X; \theta_0)'] = 0
\]

In fact, influence functions can be defined as any \( q \)-dimensional vector with mean zero, which is perpendicular to the nuisance tangent space and normalized so that

\[
E_{\theta_0}[\varphi(X)\psi_\gamma(X; \theta_0)'] = I
\]
This can be done by first finding a vector $d(X)$ which has mean zero and is perpendicular to the nuisance tangent space. Then, it can be normalized by defining

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

Alternatively, we could start off with a mean zero vector, say $a(X)$, project this vector onto the nuisance tangent space element by element and then the residuals will be orthogonal to the nuisance tangent space. Then, normalize as above. This algorithm will yield an influence function for some RAL estimator and all RAL estimators will have an influence function which can be constructed in this manner.
Example ($q = 1$)

Given a mean zero random variable, $a(X)$, we know that its projection onto the nuisance tangent space is

\[ E_{\theta_0}[a(X)\psi_\gamma(X; \theta_0)'] E_{\theta_0}[\psi_\lambda(X; \theta_0)\psi_\lambda(X; \theta_0)']^{-1} \psi_\lambda(X; \theta_0) \]

Therefore, the residual from the projection is

\[ d(X) = a(X) - E_{\theta_0}[a(X)\psi_\gamma(X; \theta_0)'] E_{\theta_0}[\psi_\lambda(X; \theta_0)\psi_\lambda(X; \theta_0)']^{-1} \psi_\lambda(X; \theta_0) \]

and after normalizing, we get

\[ \varphi(X) = E_{\theta_0}[d(X)\psi_\gamma(X; \theta_0)']^{-1} d(X) \]
Under suitable regularity conditions, we know that

$$\sqrt{n}(\hat{\theta}^{MLE}(X_n) - \theta_0) \overset{D}{\to} N(0, I^{-1}(\theta_0))$$

In which case,

$$\sqrt{n}(\hat{\gamma}^{MLE}(X_n) - \gamma_0) \overset{D}{\to} N(0, \frac{\partial\gamma(\theta_0)}{\partial\theta'} I^{-1}(\theta_0) \frac{\partial\gamma(\theta_0)}{\partial\theta})$$

Now any RAL estimator with influence function $\varphi(X)$ has an asymptotic variance equal to $E_{\theta_0}[\varphi(X)\varphi(X)']$. 
Global Efficiency of the MLE

Note that if we project \( \varphi(X) \) onto the space spanned by the entire score vector, we get the residual

\[
\varphi(X) - E_{\theta_0}[\varphi(X)\psi(X; \theta_0)']I^{-1}(\theta_0)\psi(X; \theta_0)
\]

which has variance equal to

\[
E_{\theta_0}[\varphi(X)\varphi(X)'] - E_{\theta_0}[\varphi(X)\psi(X; \theta_0)']I^{-1}(\theta_0)E_{\theta_0}[\varphi(X)\psi(X; \theta_0)']'
\]

Since \( \varphi(X) \) is an influence function of a regular estimator, we know that

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

Therefore,

\[
E_{\theta_0}[\varphi(X)\varphi(X)'] - \frac{\partial \gamma(\theta_0)}{\partial \theta'}I^{-1}(\theta_0)\frac{\partial \gamma(\theta_0)}{\partial \theta}
\]

is positive semi-definite. So,

\[
AV(\hat{\gamma}_{MLE}^n(X_n)) \leq AV(\text{any RAL estimator})
\]
The influence function for $\hat{\theta}^{MLE}(X_n)$ is given by

$$
\varphi(x) = I^{-1}(\theta_0)\psi(X; \theta_0)
$$

The influence function for $\hat{\gamma}^{MLE}(X_n) = \frac{\partial\gamma(\theta_0)}{\partial\theta'} I^{-1}(\theta_0)\psi(X; \theta_0)$. 

Consider the case, where $\theta = (\gamma, \lambda)$.

**Definition 7.3:** The *efficient score* for $\gamma$ is defined as the residual from the projection of $\psi_\gamma(X; \theta_0)$ onto $\Lambda$ (nuisance tangent space). 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 its inner product with $\psi_\gamma(X; \theta_0)$ is zero, we then define the efficient influence function

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

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

We are now in a position to prove that $\phi^\text{eff}(X)$ is the most efficient influence function.
Lemma 7:5: 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 $\Lambda \perp \{B\varphi^{\text{eff}}(X) : \text{for all } B\}$. So,

$$l(X) \in \Lambda^\perp \cap \{B\varphi^{\text{eff}}(X) : \text{for all } B\}^\perp = [\Lambda \oplus \{B\varphi^{\text{eff}}(X) : \text{for all } B\}]^\perp$$
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 $E_{\theta_0}[l(X)\psi^{\text{eff}}(X; \theta_0)'] = E_{\theta_0}[l(X)\psi(\gamma(X; \theta_0))'] = E_{\theta_0}[(\varphi(X) - \varphi^{\text{eff}}(X))\psi(\gamma(X; \theta_0))'] = 0$. Since the space spanned by $\psi^{\text{eff}}(X; \theta_0)$ 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 $\phi^{\text{eff}}(X)$ is the influence function with the smallest variance. Note that

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

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

Hence, $E_{\theta_0}[\phi(X)\phi(X)'] - E_{\theta_0}[\phi^{\text{eff}}(X)\phi^{\text{eff}}(X)]$ is positive semi-definite. This implies that $\phi^{\text{eff}}(X)$ is the most efficient influence function. The variance of the efficient influence function is $E_{\theta_0}[\psi^{\text{eff}}(X; \theta_0)\psi^{\text{eff}}(X; \theta_0)']^{-1}$, the inverse of the variance of the efficient score.
If we define $I_{\gamma\gamma}(\theta_0) = E_{\theta_0}[\psi_\gamma(X; \theta_0)\psi_\gamma(X; \theta_0)']$, $I_{\gamma\psi}(\theta_0) = E_{\theta_0}[\psi_\gamma(X; \theta_0)\psi_\lambda(X; \theta_0)']$ and $I_{\psi\psi}(\theta_0) = E_{\theta_0}[\psi_\lambda(X; \theta_0)\psi_\lambda(X; \theta_0)']$, then we get the well known result that the minimum variance for the most efficient, regular estimator is

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