Semiparametric Efficiency Bounds

Bryan S. Graham, UC - Berkeley & NBER

January 22, 2021

Let $\{Z_i\}_{i=1}^{\infty}$ be an independent and identically distributed (iid) random sequence drawn from unknown distribution function $H_0 \in \mathcal{H}$. Partition $Z = (X', Y')'$ into a vector of exogenous "instrumental" variables, X, and a vector of endogenous variables, Y. For β an unknown $K \times 1$ parameter vector of interest, and $\rho(z, \beta)$ a known $J \times 1$ vector of functions, the sole prior restriction on H_0 is that for some $\beta_0 \in \mathbb{B}$ (with $\mathbb{B}$ and open subset of $\mathbb{R}^K$) and all $x \in \mathbb{X}$

$$\mathbb{E}\left[\rho(Z, \beta_0)| X = x\right] = 0. \tag{1}$$

Our *semiparametric model* consists of the iid sampling assumption, restriction (1) and various regularity conditions.

Using the methods of Bickel et al. (1993), this note shows how to derive a lower bound on the asymptotic sampling variance of any regular asymptotically linear (RAL) estimate of β_0. In a beautiful paper, Chamberlain (1987) also considers this problem using multinomial approximation arguments. I will not cover the Chamberlain (1987) method here, but it is worth studying. I have also found it useful in my own work (e.g., Graham, 2011).

The development here is informal, as well as skeletal, but it is sufficiently detailed to convey key ideas and to provide a template for how to go about efficiency bound calculations in practice. We'll apply the tools developed here in our examination of missing data problems later in the course.

Parametric sub-model

The basic idea behind semiparametric efficiency bound (SEB) calculations comes from a thought experiment due to Stein (1956).[1] Assume that $Z_1, Z_2, \ldots$ are independent random

[1]Stein worked with Ken Arrow, among others, in the meteorology department at the Pentagon during World War II. Evidently his thesis topic was, at least in part, due to a suggestion from Arrow. Stein was

draws from some parametric distribution with density $h(z,\eta)$. Here $h(z,\eta)$ corresponds to a *parametric sub-model* indexed by the parameter η. Let $h_0(z)$ denote the true distribution of Z. We will assume that when $\eta = \eta_0$ we have that $h(z,\eta_0) = h_0(z)$: our parametric sub-model contains the truth. The "sub" prefix to model highlights the fact that $h(z,\eta_0)$ is a subset of the larger model consisting of *all* distributions which satisfy the restrictions imposed by the semiparametric model.

Since (1) imposes no restrictions on the marginal distribution of X, we will only consider parametric sub-models for the conditional distribution of Y given $X = x$ (cf., Newey, 2004). An example of such a parametric sub-model is the exponential family

$$h(y|\, x;\eta) = h_0(y|\, x)\exp\left(t(y)'\mu(x,\eta) - \phi(x,\eta)\right),$$

with $t(y)$ a vector of linearly independent functions of y and

$$\phi(x,\eta) = \ln\left[\int h_0(y|\, x)\exp\left(t(y)'\mu(x,\eta)\right)\mathrm{d}y\right].$$

Here $\mu(x,\eta)$ is a known function of x and η that equals zero at $\eta = \eta_0$; this yields $h(y|\, x;\eta_0) = h_0(y|\, x)$ as required. See Back & Brown (1992).

Observe that the parameter of interest, β, is a function $\beta(\eta)$ of the sub-model parameter η:

$$\int \rho(z,\beta(\eta_0))\, h(y|\, x,\eta_0)\,\mathrm{d}z = 0. \tag{2}$$

Our parametric sub-model (i) satisfies the semiparametric restriction (1) and (ii) contains the truth. Henceforth these conditions will be assumed to hold when referring to a parametric sub-model.

Let $\mathcal{I}(\eta_0) = \mathbb{E}\left[\mathbb{S}_\eta \mathbb{S}_\eta'\right]$ be the Fisher information matrix with $\mathbb{S}_\eta = \left.\frac{\partial \ln h(Y|X,\eta)}{\partial \eta}\right|_{\eta=\eta_0}$. The (conditional) MLE of η_0 is asymptotically normal with

$$\sqrt{N}(\hat{\eta} - \eta) \xrightarrow{D} N\left(0, \mathcal{I}^{-1}(\eta_0)\right).$$

By the Delta-method, as well as the invariance properties of maximum likelihood, the Cramer-Rao (CR) lower bound for β_0 is therefore

$$\frac{\partial \beta(\eta)}{\partial \eta'}\mathbb{E}\left[\mathbb{S}_\eta \mathbb{S}_\eta'\right]\left[\frac{\partial \beta(\eta)}{\partial \eta'}\right]'. \tag{3}$$

in the Berkeley Statistics Department for a number of years. He moved to Stanford in order to avoid the "loyalty oaths" of the McCarthy period.

Note that our parametric sub-model, since it is written in terms of the true conditional density of Y given X, $h_0(y|x)$, does not offer a practical method of estimating β_0. We are simply performing a thought experiment: how precisely could we estimate β_0 in large samples if we knew the data were generated according to this particular parametric family?

At this point, picking up the argument of Stein (1956), observe that any consistent and asymptotically normal semi-parametric estimator must have a sampling variance at least as large as the CR lower bound associated with our sub-model. Since this is true for any sub-model, we conclude that the asymptotic variance of any semiparametric estimator is bounded below by the supremum of the CR lower bounds across all parametric sub-models.

As will become clear, this thought experiment will also help us calculate the semiparametric efficiency bound.

Tangent space

There are many possible parametric sub-models consistent with our semiparametric model. It will be useful to have a characterization of the entire *set* of such models. Such a characterization is provided by the *tangent set.* The model tangent set $\mathcal{T}$ consists of the closed linear span of all possible parametric sub-model scores. To characterize the tangent set we need to understand the types of scores "allowed" by our model.

By the usual conditional mean zeroness of the score function, we know that $\mathbb{E}\left[\left.\mathbb{S}_\eta\right| X\right]=0$. So any element of $\mathcal{T}$ must be conditionally mean zero given X. However restriction (1) imposes additional restrictions on allowable scores. To see the form of these restrictions differentiate (2) with respect to η to get

$$\int \frac{\partial \rho\left(z,\beta_0\right)}{\partial \beta'}\frac{\partial \beta\left(\eta\right)}{\partial \eta'}h\left(\left.y\right|x,\eta_0\right)\mathrm{d}z+\int \rho\left(z,\beta_0\right)\frac{\partial h\left(\left.y\right|x,\eta_0\right)}{\partial \eta'}\mathrm{d}z=0.$$

Defining $\Gamma_0\left(x\right)=\mathbb{E}\left[\left.\frac{\partial \rho\left(Z,\beta_0\right)}{\partial \beta'}\right| X=x\right]$ and re-arranging gives the equality

$$-\underset{J\times K}{\Gamma_0\left(X\right)}\underset{K\times L}{\frac{\partial \beta\left(\eta\right)}{\partial \eta'}}=\mathbb{E}\left[\left.\underset{J\times 1}{\rho\left(Z,\beta_0\right)}\underset{1\times L}{\mathbb{S}_\eta'}\right| X\right] \tag{4}$$

with $K=\dim\left(\beta\right)$, $J=\dim\left(\rho\left(Z,\beta_0\right)\right)$ and $L=\dim\left(\eta\right)$

Restriction (1) implies a structured form of covariance between any allowable score vector and $\rho(Z, \beta_0)$. Putting things together suggests a tangent set of

$$\begin{aligned} \mathcal{T} = \Big\{ & s(Z) \, : \, \mathbb{E}\left[s(Z)^2 \right] < \infty, \mathbb{E}\left[s(Z) | X \right] = 0 \, , \\ & \mathbb{E}\left[\rho(Z, \beta_0) s(Z) | X \right] = \Gamma_0(X) c \text{ for a constant } K \times 1 \text{ vector } c \Big\} . \end{aligned} \tag{5}$$

Formally (5) is just a conjecture on the form of $\mathcal{T}$. As the example illustrates, there is some "art" involved in constructing a correct guess about the form of the tangent space. Newey (1990a, Appendix B) discusses how to formally verify tangent set conjectures. This is a technical exercise and can often be skipped in practice. See also Tsiatis (2006).

A GMM estimator

The previous two sections respectively introduced (i) an infeasible maximum likelihood estimate of β_0 and (ii) characterized the class of all possible parametric sub-models allowed by the semiparametric model. Here I introduce a consistent and asymptotically normal semiparametric estimate of β_0. These properties – consistency and normality – hold across all parametric sub-models allowed by the semiparametric model. This model will help us derive an alternative representation of the CR lower bound, equation (3) above. This alternative representation will, in turn, both motivate a (more) formal definition of the semiparametric efficiency bound as well as suggest a method for calculating it.

Let $A(x, \beta)$ be a $K \times J$ instrument matrix. Our semiparametric estimate of β_0 is the solution to

$$\frac{1}{N} \sum_{i=1}^{N} A\left(X_i, \hat{\beta} \right) \rho\left(Z_i, \hat{\beta} \right) = 0. \tag{6}$$

Let $A_k(x, \beta)$ denote the k^{th} row of $A(x, \beta)$ and

$$\underset{K \times K}{B(Z_i, \beta)} \overset{def}{\equiv} \left[\begin{array}{c} \left\{ \frac{\partial A_1(X_i, \beta)}{\partial \beta} \rho(Z_i, \beta) \right\}' \\ \vdots \\ \left\{ \frac{\partial A_K(X_i, \beta)}{\partial \beta} \rho(Z_i, \beta) \right\}' \end{array} \right].$$

A mean value expansion of (6) gives

$$
\begin{aligned}
0 =& \frac{1}{N}\sum_{i=1}^{N} A\left(X_i,\beta_0\right)\rho\left(Z_i,\beta_0\right) = 0 \\
&+ \frac{1}{N}\sum_{i=1}^{N}\left\{B\left(Z_i,\bar{\beta}\right) + A\left(X_i,\bar{\beta}\right)\frac{\partial\rho\left(Z_i,\bar{\beta}\right)}{\partial\beta'}\right\}\left(\hat{\beta}-\beta_0\right).
\end{aligned}
$$

By (1), the LLN, and iterated expectations $\frac{1}{N}\sum_{i=1}^{N} B\left(Z_i,\bar{\beta}\right) \xrightarrow{p} 0$ and $\frac{1}{N}\sum_{i=1}^{N} A\left(X_i,\bar{\beta}\right)\frac{\partial\rho\left(Z_i,\bar{\beta}\right)}{\partial\beta'} \xrightarrow{p}$ $\mathbb{E}\left[A_0\left(X\right)\Gamma_0\left(X\right)\right]$ for $A_0\left(x\right) = A\left(x,\beta_0\right)$ (assuming, as I do, that $\hat{\beta} \xrightarrow{p} \beta_0$). Standard arguments then yield the asymptotically linear representation

$$
\sqrt{\mathrm{N}}\left(\hat{\beta}-\beta_0\right) = -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\underbrace{\mathbb{E}\left[A_0\left(X\right)\Gamma_0\left(X\right)\right]^{-1}\times A\left(X_i,\beta_0\right)\rho\left(Z_i,\beta_0\right)}_{\phi\left(Z_i,\beta_0\right)} + o_p\left(1\right). \tag{7}
$$

It is convenient to write the influence function, $\phi\left(Z_i,\beta_0\right)$, in the "M-estimator" form

$$
\phi\left(Z_i,\beta_0\right) = -M_0^{-1}m\left(Z_i,\beta_0\right), \tag{8}
$$

with $M_0 = \mathbb{E}\left[A_0\left(X\right)\Gamma_0\left(X\right)\right]$ and $m\left(Z_i,\beta_0\right) = A\left(X_i,\beta_0\right)\rho\left(Z_i,\beta_0\right)$. From (8) it follows that

$$
\sqrt{\mathrm{N}}\left(\hat{\beta}-\beta_0\right) \xrightarrow{D} \mathcal{N}\left(0, M_0^{-1}\mathbb{E}\left[mm'\right]\left(M_0^{-1}\right)'\right).
$$

Semiparametric Efficiency Bound

Return to our parametric sub-model setting. From (1) we have that, for $m\left(z,\beta\right)$ as defined above,

$$
\mathbb{E}\left[m\left(Z,\beta\left(\eta_0\right)\right)\right] = \int m\left(z,\beta\left(\eta_0\right)\right)h\left(y|\,x,\eta_0\right)h\left(x,\eta_0\right)\mathrm{d}z = 0.
$$

Differentiating with respect to β yields

$$
\begin{aligned}
0 =& \int \frac{\partial m\left(z,\beta_0\right)}{\partial\beta'}\frac{\partial\beta\left(\eta_0\right)}{\partial\eta'}h\left(y|\,x,\eta_0\right)h\left(x,\eta_0\right)\mathrm{d}z \\
&+ \int m\left(z,\beta_0\right)\frac{\partial h\left(y|\,x,\eta_0\right)}{\partial\eta'}h\left(x,\eta_0\right)\mathrm{d}z \\
&+ \int m\left(z,\beta_0\right)h\left(y|\,x,\eta_0\right)\frac{\partial h\left(x,\eta_0\right)}{\partial\eta'}\mathrm{d}z.
\end{aligned}
$$

Rearrangement then gives

$$\begin{aligned}\frac{\partial\beta\left(\eta_{0}\right)}{\partial\eta'} &= -M_{0}^{-1}\mathbb{E}\left[m\left(Z,\beta_{0}\right)\left\{\mathbb{S}_{\eta}+\mathbb{T}_{\eta}\right\}'\right] \\ &= -M_{0}^{-1}\mathbb{E}\left[m\left(Z,\beta_{0}\right)\mathbb{S}_{\eta}'\right], \end{aligned} \tag{9}$$

with $\mathbb{T}_{\eta}=\left.\frac{\partial\ln h(X,\eta)}{\partial\eta}\right|_{\eta=\eta_{0}}$ and the second equality an implication of (1). Equation (9) corresponds to the Generalized Information Matrix Equality (GIME). See Newey (1990b, p. 104).

Equations (3) and (9) yield the alternative representation of the Cramer-Rao (CR) lower bound for β_0:

$$\mathcal{I}_{\eta}^{-1}\left(\beta_{0}\right)=M_{0}^{-1}\mathbb{E}\left[m\mathbb{S}_{\eta}'\right]\mathbb{E}\left[\mathbb{S}_{\eta}\mathbb{S}_{\eta}'\right]^{-1}\mathbb{E}\left[\mathbb{S}_{\eta}m'\right]\left(M_{0}^{-1}\right)'. \tag{10}$$

We can use this form of the CR lower bound to answer the following question: how efficient is our semiparametric estimate of β_0 relative to the (infeasible) MLE based upon a particular parametric sub-model? From the previous section we have an asymptotic variance for our semiparametric estimate of

$$\mathrm{AVar}\left(\sqrt{\mathrm{N}}\left(\hat{\beta}-\beta_{0}\right)\right)=\mathbb{E}\left[M_{0}^{-1}mm'\left(M_{0}^{-1}\right)'\right].$$

This, and our alternate form for the CR lower bound (10), yield after some rearrangement

$$\begin{aligned}\mathrm{AVar}\left(\sqrt{\mathrm{N}}\left(\hat{\beta}-\beta_{0}\right)\right)-\mathcal{I}_{\eta}^{-1}\left(\beta_{0}\right)=&M_{0}^{-1}\left[\left(m-\mathbb{E}\left[m\mathbb{S}_{\eta}'\right]\mathbb{E}\left[\mathbb{S}_{\eta}\mathbb{S}_{\eta}'\right]^{-1}\mathbb{S}_{\eta}\right)\right. \\ &\left.\left(m-\mathbb{E}\left[m\mathbb{S}_{\eta}'\right]\mathbb{E}\left[\mathbb{S}_{\eta}\mathbb{S}_{\eta}'\right]^{-1}\mathbb{S}_{\eta}\right)'\right]\left(M_{0}^{-1}\right)'.\end{aligned} \tag{11}$$

A few observations regarding (11):

1. Since the difference (11) is positive semi-definite our semiparametric estimate is never (asymptotically) more precise than the appropriate – but infeasible – MLE.

2. Since (11) holds for *any* parametric sub-model it necessarily holds vis-a-vis the parametric sub-model with the highest asymptotic variance. This means that our semiparametric estimate will never have an asymptotic variance smaller than the supremum of CR bounds across all parametric sub-models.

These observations apply to *any* regular asymptotically linear (RAL) semiparametric estimator (see equation (7) and Newey (1990b)). Hence the supremum across all possible

CR lower bounds provides a notion of semiparametric efficiency – it is not possible for a semiparametric estimator to "do better" than this supremum.

Two questions remain:

1. How do we calculate the semiparametric variance bound?

2. How do we construct feasible estimators which attain this bound?

I only address the first question here. See Newey (1990b) for a discussion of the second as well as the examples we'll cover later in this course.

Calculating the bound

The form of (11) provides a hint as to how to calculate the semiparametric variance bound. Observe that the CR lower bound (10) corresponds to the variance of

$$\delta_{\eta}\left(Z,\beta_{0}\right)=\mathbb{E}\left[M_{0}^{-1}m\mathbb{S}_{\eta}^{\prime}\right]\mathbb{E}\left[\mathbb{S}_{\eta}\mathbb{S}_{\eta}^{\prime}\right]^{-1}\mathbb{S}_{\eta},$$

which equals the prediction from a population linear regression of the influence function $\phi\left(Z,\beta_{0}\right)=-M_{0}^{-1}m\left(Z,\beta_{0}\right)$ onto the parametric sub-model scores, $\mathbb{S}_{\eta}$. To find the variance bound we need to maximize the variance of these predictions. Our maximization is over the set of all possible parametric sub-model scores allowed by the semiparametric model. Maximizing predictor variance is equivalent to minimizing the variance of the population residual:

$$\phi\left(Z,\beta_{0}\right)-\delta_{\eta}\left(Z,\beta_{0}\right).$$

Imagine regressing $M_{0}^{-1}m\left(z,\beta_{0}\right)$ onto an increasing set of scores allowed by the semiparametric model. Adding scores in this way can only reduce residual variance. The set of all possible scores is given by the tangent set $\mathcal{T}$. Finding the semiparametric efficiency bound therefore corresponds to calculating the *projection* of $\phi\left(Z,\beta_{0}\right)$ onto the tangent set. Depending on the problem calculating this projection may be straightforward or very hard. Newey (1990a) provides several examples (see also Appendix B.10.3 of Bickel & Doksum (2015) and Tsiatis (2006)). For more details on projections see the Projection Theorem lecture notes.

In many cases a "guess and check" approach is easiest. This involves guessing the form of the required projection and then verifying that is satisfies the conditions of the Projection Theorem. See Theorem 3.1 of Newey (1990b) and the accompanying discussion. Graham (2011) provides an example of the guess and check approach for a relatively complicated model.

Let $\delta_{\text{eff}}(Z,\beta_0)$ be the required projection $\Pi\left(\phi(Z,\beta_0)|\,\mathcal{T}\right)$; this projection is called the *efficient influence function*. The variance of the efficient influence function coincides with the semiparametric variance bound:

$$\mathcal{I}^{-1}(\beta_0) = \mathbb{E}\left[\delta_{\text{eff}}(Z,\beta_0)\,\delta_{\text{eff}}(Z,\beta_0)'\right]. \tag{12}$$

Any semiparametrically efficient estimator will be asymptotically linear with:

$$\sqrt{N}\left(\hat{\beta}-\beta_0\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\delta_{\text{eff}}(Z_i,\beta_0) + o_p(1).$$

Hence the terminology "efficient influence function".

Efficiency bound for conditional moment problem

For our conditional moment problem inspection of the parametric sub-model score restriction (4) suggests the following (infeasible) estimate might be efficient:

$$\frac{1}{N}\sum_{i=1}^{N}\Gamma_0(X_i)'\,\Omega_0^{-1}(X_i)\,\rho\left(Z_i,\hat{\beta}\right) = 0,$$

where $\Omega_0(X) = \mathbb{E}\left[\rho(Z,\beta_0)\,\rho(Z,\beta_0)'\middle|\,X\right]$ equals the conditional variance of $\rho(Z,\beta_0)$ given X. By the usual arguments, this infeasible estimate is asymptotically linear:

$$\sqrt{N}\left(\hat{\beta}-\beta\right) = -\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\Gamma_0(X)'\,\Omega_0^{-1}(X)\,\Gamma_0(X)\right]^{-1}\Gamma_0(X_i)'\,\Omega_0^{-1}(X_i)\,\rho(Z_i,\beta_0) + o_p(1).$$

We therefore conjecture that the efficient influence function is

$$\delta_{\text{eff}}(Z_i,\beta_0) = \mathbb{E}\left[\Gamma_0(X)'\,\Omega_0^{-1}(X)\,\Gamma_0(X)\right]^{-1}\Gamma_0(X_i)'\,\Omega_0^{-1}(X)\,\rho(Z_i,\beta_0).$$

Let $\rho = \rho(Z,\beta_0)$ and $m = m(Z,\beta_0)$. To verify this conjecture we check the orthogonality condition of Theorem 3.1 of Newey (1990b):

$$\mathbb{E}\left[\left\{-M_0^{-1}m + \mathbb{E}\left[\Gamma_0(X)'\,\Omega_0^{-1}(X)\,\Gamma_0(X)\right]^{-1}\Gamma_0(X)'\,\Omega_0^{-1}(X)\,\rho\right\}s(Z)\right] = 0,\ \ s(Z)\in\mathcal{T}.$$

Recall that $m(Z,\beta_0) = A_0(X)\rho(Z,\beta_0)$ and $M_0 = \mathbb{E}[A_0(X)\Gamma_0(X)]$, manipulation, iterated expectations, and the fact that $\mathbb{E}[\rho(Z,\beta_0)s(Z)|X] = \Gamma_0(X)c$ gives

$$
\begin{aligned}
0 &= \mathbb{E}\left[\left\{-M_0^{-1}m + \mathbb{E}\left[\Gamma_0(X)'\Omega_0^{-1}(X)\Gamma_0(X)\right]^{-1}\Gamma_0(X)'\Omega_0^{-1}(X)\rho\right\}s(Z)\right] \\
&= \mathbb{E}\left[\left\{-M_0^{-1}A_0(X) + \mathbb{E}\left[\Gamma_0(X)'\Omega_0^{-1}(X)\Gamma_0(X)\right]^{-1}\Gamma_0(X)'\Omega_0^{-1}(X)\right\}\rho s(Z)\right] \\
&= \mathbb{E}\left[\left\{-M_0^{-1}A_0(X) + \mathbb{E}\left[\Gamma_0(X)'\Omega_0^{-1}(X)\Gamma_0(X)\right]^{-1}\Gamma_0(X)'\Omega_0^{-1}(X)\right\}\mathbb{E}[\rho s(Z)|X]\right] \\
&= \mathbb{E}\left[\left\{-M_0^{-1}A_0(X) + \mathbb{E}\left[\Gamma_0(X)'\Omega_0^{-1}(X)\Gamma_0(X)\right]^{-1}\Gamma_0(X)'\Omega_0^{-1}(X)\right\}\Gamma_0(X)\right]c \\
&= -\mathbb{E}[A_0(X)\Gamma_0(X)]^{-1}\mathbb{E}[A_0(X)\Gamma_0(X_i)]c \\
&+ \mathbb{E}\left[\Gamma_0(X)'\Omega_0^{-1}(X)\Gamma_0(X)\right]^{-1}\mathbb{E}\left[\Gamma_0(X)'\Omega_0^{-1}(X)\Gamma_0(X)\right]c \\
&= -I_K c + I_K c = 0
\end{aligned}
$$

as required.

The semiparametric efficiency bound is therefore

$$
\mathcal{I}^{-1}(\beta_0) = \mathbb{E}\left[\Gamma_0(X)'\Omega_0^{-1}(X)\Gamma_0(X)\right].
$$

See Chamberlain (1987) for an alternative derivation of this bound.

Further reading

Newey (1990b) provides an excellent, if now somewhat dated, introductory survey of semi-parametric efficiency bounds. Tsiatis (2006) is a useful reference with more details. Bickel et al. (1993) is the seminal reference; this book is very challenging. Severini & Tripathi (2013) provide a more recent survey, with many examples.

References

Back, K. & Brown, D. P. (1992). Gmm, maximum likelihood, and nonparametric efficiency. *Economic Letters*, 39(1), 23 – 28.

Bickel, P. J. & Doksum, K. A. (2015). *Mathematical Statistics*, volume 1. Boca Raton: Chapman & Hall, 2nd edition.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., & Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. New York: Springer-Verlag.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3), 305 – 334.

Graham, B. S. (2011). Efficiency bounds for missing data models with semiparametric restrictions. *Econometrica*, 79(2), 437 – 452.

Newey, W. K. (1990a). Efficient instrumental variables estimation of nonlinear models. *Econometrica 58*, 58(4), 809 – 837.

Newey, W. K. (1990b). Semiparametric efficiency bounds. *Journal of Applied Econometrics*, 5(2), 99 – 135.

Newey, W. K. (2004). Efficient semiparametric estimation via moment restrictions. *Econometrica*, 72(6), 1877 – 1897.

Severini, T. A. & Tripathi, G. (2013). Semiparametric efficiency bounds for microeconometric models: a survey. *Foundations and Trends in Econometrics*, 6(3-4), 163 – 397.

Stein, C. M. (1956). Efficient nonparametric testing and estimation. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1(187 - 195).

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. New York: Springer.

