ECONOMETRICA
JOURNAL OF THE ECONOMETRIC SOCIETY
An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics
http://www.econometricsociety.org/

Econometrica, Vol. 79, No. 2 (March, 2011), 437–452

EFFICIENCY BOUNDS FOR MISSING DATA MODELS WITH SEMIPARAMETRIC RESTRICTIONS

BRYAN S. GRAHAM
New York University, New York, NY 10012, U.S.A. and National Bureau of Economic Research

---

Econometrica, Vol. 79, No. 2 (March, 2011), 437–452

EFFICIENCY BOUNDS FOR MISSING DATA MODELS WITH SEMIPARAMETRIC RESTRICTIONS

BY BRYAN S. GRAHAM[1]

This paper shows that the semiparametric efficiency bound for a parameter identified by an unconditional moment restriction with data missing at random (MAR) coincides with that of a particular augmented moment condition problem. The augmented system consists of the inverse probability weighted (IPW) original moment restriction and an additional conditional moment restriction which exhausts all other implications of the MAR assumption. The paper also investigates the value of additional semiparametric restrictions on the conditional expectation function (CEF) of the original moment function given always observed covariates. In the program evaluation context, for example, such restrictions are implied by semiparametric models for the potential outcome CEFs given baseline covariates. The efficiency bound associated with this model is shown to also coincide with that of a particular moment condition problem. Some implications of these results for estimation are briefly discussed.

KEYWORDS: Missing data, semiparametric efficiency, propensity score, (augmented) inverse probability weighting, double robustness, average treatment effects, causal inference.

1. INTRODUCTION

LET $Z = (Y_1', X')'$ BE A VECTOR of modelling variables, let $\{Z_i\}_{i=1}^{\infty}$ be an independent and identically distributed random sequence drawn from the unknown distribution F_0, let β be a $K \times 1$ unknown parameter vector, and let $\psi(Z, \beta)$ be a known vector-valued function of the same dimension.[2] The only prior restriction on F_0 is that for some $\beta_0 \in \mathcal{B} \subset \mathbb{R}^K$,

$$\mathbb{E}[\psi(Z, \beta_0)] = 0. \tag{1}$$

Chamberlain (1987) showed that the maximal asymptotic precision with which β_0 can be estimated under (1) (subject to identification and regularity con-

[1]I would like to thank Gary Chamberlain, Jinyong Hahn, Guido Imbens, Michael Jansson, and Whitney Newey for comments on earlier draft. Helpful discussions with Oliver Linton, Cristine Pinto, Jim Powell, and Geert Ridder as well as participants in the Berkeley Econometrics Reading Group and Seminars are gratefully acknowledged. This revision has benefited from Tom Rothenberg's skepticism, discussions with Michael Jansson, Justin McCrary, Jim Powell, and the comments of a co-editor and three especially meticulous/generous anonymous referees. All the usual disclaimers apply. This is a heavily revised version of material that previously circulated under the titles "A Note on Semiparametric Efficiency in Moment Condition Models With Missing Data," "GMM 'Equivalence' for Semiparametric Missing Data Models," and "Efficient Estimation of Missing Data Models Using Moment Conditions and Semiparametric Restrictions."

[2]Extending what follows to the overidentified case is straightforward.

 DOI: 10.3982/ECTA7379

ditions) is given by $\mathcal{I}_{\mathrm{f}}(\beta_0) = \Gamma_0' \Omega_0^{-1} \Gamma_0$, with $\Gamma_0 = \mathbb{E}[\partial\psi(Z, \beta_0)/\partial\beta']$ and $\Omega_0 = \mathbb{V}(\psi(Z, \beta_0))$.[3]

Now consider the case where a random sequence from F_0 is unavailable; instead, only a selected sequence of samples is available. Let D be a binary selection indicator. When $D = 1$, we observe Y_1 and X; when $D = 0$, we observe only X.[4] This paper considers estimation of β_0 under restriction (1) and the following additional assumptions.

ASSUMPTION 1.1—Random Sampling: *$\{Z_i, D_i\}_{i=1}^{\infty}$ is an independent and identically distributed random sequence from F_0.*

ASSUMPTION 1.2—Observed Data: *For each unit, we observe D, X, and $Y = DY_1$.*

ASSUMPTION 1.3—Conditional Independence: $Y_1 \perp D | X$.

ASSUMPTION 1.4—Overlap: $0 < \kappa \leq p_0(x) \leq 1$ *for* $p_0(x) = \Pr(D = 1| X = x)$ *and for all* $x \in \mathcal{X} \subset \mathbb{R}^{\dim(x)}$.

Restriction (1) and Assumptions 1.1–1.4 constitute a semiparametric model for the data. Henceforth, I refer to this model as the semiparametric missing data model or the missing at random (MAR) setup. Robins, Rotnitzky, and Zhao (1994, Proposition 2.3, p. 850) derived the efficient influence function for this problem and proposed a locally efficient augmented inverse probability weighting (AIPW) estimator (cf. Scharfstein, Rotnitzky, and Robins (1999), Bang and Robins (2005), Tsiatis (2006)). Cheng (1994), Hahn (1998), Hirano, Imbens, and Ridder (2003), Imbens, Newey, and Ridder (2005), and Chen, Hong, and Tarozzi (2008) developed globally efficient estimators.

The MAR setup has been applied to a number of important econometric and statistical problems, including program evaluation as surveyed by Imbens (2004), nonclassical measurement error (e.g., Robins, Hsieh, and Newey (1995), Chen, Hong, and Tamer (2005)), missing regressors (e.g., Robins, Rotnitzky, and Zhao (1994)), attrition in panel data (e.g., Robins, Rotnitzky, and Zhao (1995), Robins and Rotnitzky (1995), Wooldridge (2002)), and M-estimation under variable probability sampling (e.g., Wooldridge (1999a,

[3]Throughout uppercase letters denote random variables, lowercase letters denote specific realizations of them, and calligraphic letters denote their support. I use the notation $\mathbb{E}[A|c] = \mathbb{E}[A|C = c]$, $\mathbb{V}(A|c) = \mathrm{Var}(A|C = c)$, and $\mathbb{C}(A, B|c) = \mathrm{Cov}(A, B|C = c)$.

[4]An earlier version of this paper considered the slightly more general setup with $\psi(Z, \beta) = \psi_1(Y_1, X, \beta) - \psi_0(Y_0, X, \beta)$ with (X, Y) observed, where $Y = DY_1 + (1 - D)Y_0$. Results for this extended model, which contains the standard causal inference model and the two-sample instrumental variables model as special cases (cf. Imbens (2004), Angrist and Krueger (1992)), follow directly and straightforwardly from those outlined below.

2007)). Chen, Hong, and Tarozzi (2004), Wooldridge (2007), and Graham, Pinto, and Egel (2010) discussed several other applications.

The maximal asymptotic precision with which β_0 can be estimated under the MAR setup has been characterized by Robins, Rotnitzky, and Zhao (1994) and is given by

$$\mathcal{I}_{\mathrm{m}}(\beta_0) = \Gamma_0' \Lambda_0^{-1} \Gamma_0, \tag{2}$$

with $\Lambda_0 = \mathbb{E}[\Sigma_0(X)/p_0(X) + q(X;\beta_0)q(X;\beta_0)']$, where $\Sigma_0(x) = \mathbb{V}(\psi(Z,\beta_0)|x)$ and $q(x;\beta) = \mathbb{E}[\psi(Z,\beta)|x]$.

The associated efficient influence function, also due to Robins, Rotnitzky, and Zhao (1994), is given by

$$\phi(z,\theta_0) = \Gamma_0^{-1} \times \left\{ \frac{d}{p_0(x)} \psi(z,\beta_0) - \frac{q(x;\beta_0)}{p_0(x)}(d - p_0(x)) \right\} \tag{3}$$

for $\theta = (p, q', \beta')'$.

The calculation of (2) is now standard. Knowledge of (2) is useful because it quantifies the cost—in terms of asymptotic precision—of the missing data and because it can be used to verify whether a specific estimator for β_0 is efficient. To simplify what follows, I will explicitly assume that $\mathcal{I}_{\mathrm{m}}(\beta_0)$ is well defined (i.e., that all its component expectations exist and are finite, and that all its component matrices are nonsingular).

This paper shows that the semiparametric efficiency bound for β_0 under the MAR setup coincides with the bound for a particular augmented moment condition problem. The augmented system consists of the inverse probability of observation weighted (IPW) original moment restriction (1) and an additional conditional moment restriction that exhausts all other implications of the MAR setup. This general equivalence result, while implicit in the form of the efficient influence function (3), is apparently new. It provides fresh intuitions for several "paradoxes" in the missing data literature, including the well known results that projection onto, or weighting by the inverse of, a known propensity score results in inefficient estimates (e.g., Hahn (1998), Hirano, Imbens, and Ridder (2003)), that smoothness and exclusion priors on the propensity score do not increase the precision with which β_0 can be estimated (Robins, Hsieh, and Newey (1995), Robins and Rotnitzky (1995), Hahn (1998, 2004)), and that weighting by a nonparametric estimate of the propensity score results in an efficient estimator (Hirano, Imbens, and Ridder (2003); cf. Hahn (1998), Wooldridge (2007), Prokhorov and Schmidt (2009), Hitomo, Nishiyama, and Okui (2008)).

This paper also analyzes the effect of imposing additional semiparametric restrictions on the conditional expectation function (CEF) $q(x;\beta) = \mathbb{E}[\psi(Z,\beta)|x]$. If $\psi(Z,\beta) = Y_1 - \beta$, as when the target parameter is $\beta_0 = \mathbb{E}[Y_1]$, then such restrictions may arise from prior information on the form of $\mathbb{E}[Y_1|x]$. Such restrictions may arise in other settings as well. For example, if the goal is to estimate a vector of linear predictor coefficients in the presence of missing regressors, then a semiparametric model for the CEFs of the missing regressors given always observed variables generates restrictions on the form of $q(x; \beta)$ (cf. Robins, Rotnitzky, and Zhao (1994)).[5]

Formally I consider the semiparametric model defined by restriction (1), Assumptions 1.1–1.4 and the following additional assumption.

ASSUMPTION 1.5—Functional Restriction: *For* $X = (X_1', X_2')'$ *let*

$$\mathbb{E}[\psi(Z, \beta_0)|x] = q(x, \delta_0, h_0(x_2); \beta_0),$$

where $q(x, \delta, h(x_2); \beta)$ *is a known* $K \times 1$ *function*, δ *is a* $J \times 1$ *finite-dimensional unknown parameter*, *and* $h(\cdot)$ *is an unknown function mapping from a subset of* $\mathcal{X}_2 \subset \mathbb{R}^{\dim(X_2)}$ *into* $\mathcal{H} \subset \mathbb{R}^P$.

To the best of my knowledge, the variance bound for this problem—the MAR setup with "functional" restrictions—has not been previously calculated. In an innovative paper, Wang, Linton, and Härdle (2004) considered a special case of this model where $\psi(Z, \beta) = Y_1 - \beta$. They imposed a partial linear structure, as in Engle, Granger, Rice, and Weiss (1986), on $\mathbb{E}[Y_1|x]$ such that $q(x, \delta_0, h_0(x_2); \beta_0) = x_1'\delta_0 + h_0(x_2) - \beta_0$. In making their variance bound calculation, they assumed that the conditional distribution of Y_1 given X is normal with a variance that does not depend on X. They did not provide a bound for the general case, but conjectured that it is "very complicated" (Wang, Linton, and Härdle (2004, p. 338)). The result given below extends their work to moment condition models, general forms for $q(x, \delta, h(x_2); \beta)$, and, importantly, does not require that $\psi(Z, \beta)$ be conditionally normally distributed and/or homoscedastic.

Augmenting the MAR setup with Assumption 1.5 generates a middle ground between the fully parametric likelihood-based approaches to missing data described by Little and Rubin (2002) and those which leave $\mathbb{E}[\psi(Z, \beta_0)|x]$ unrestricted (e.g., Cheng (1994), Hahn (1998), Hirano, Imbens, and Ridder (2003)). Likelihood-based approaches are very sensitive to misspecification (cf. Imbens (2004)), while approaches which utilize only the basic MAR setup require high dimensional smoothing which may deleteriously affect small sample performance (cf. Wang, Linton, and Härdle (2004), Ichimura and Linton (2005)). Assumption 1.5 is generally weaker than a parametric specification for the conditional distribution of $\psi(Z, \beta_0)$ given X, but at the same time reduces the dimension of the nonparametric smoothing problem. Below I show how to efficiently exploit prior information on the form of $\mathbb{E}[\psi(Z, \beta_0)|x]$. I also provide conditions under which consistent estimation of β_0 is possible even if the exploited information is incorrect.

[5]The formation of predictive models of this type is the foundation of the imputation approach to missing data described by Little and Rubin (2002).

Section 2 reports the first result of the paper: an equivalence between the MAR setup and a particular method-of-moments problem. Equivalence, which is suggested by the form of the efficient influence function derived by Robins, Rotnitzky, and Zhao (1994), was previously noted for special cases by Newey (1994a) and Hirano, Imbens, and Ridder (2003). I discuss the connection between their results and the general result provided below. I also highlight some implications of the equivalence result for understanding various aspects of the MAR setup. Section 3 calculates the variance bound for β_0 when the MAR setup is augmented by Assumption 1.5. I discuss when Assumption 1.5 is likely to be informative and also when consistent estimation is possible even if it is erroneously maintained.

2. EQUIVALENCE RESULT

Under the MAR setup, the inverse probability weighted (IPW) moment condition

$$\mathbb{E}\left[\frac{D}{p_0(X)}\psi(Z,\beta_0)\right]=0 \tag{4}$$

is valid (e.g., Hirano, Imbens, and Ridder (2003), Wooldridge (2007)). The conditional moment restriction

$$\mathbb{E}\left[\frac{D}{p_0(X)}-1\middle|X\right]=0 \quad \forall X\in\mathcal{X} \tag{5}$$

also holds and nonparametrically identifies $p_0(x)$. While the terminology is inexact, in what follows I call (4) the *identifying moment* and (5) the *auxiliary moment.*

Consider the case where $p_0(x)$ is known such that (5) is truly an auxiliary moment. One efficient way to exploit the information (5) contains is, following Newey (1994a) and Brown and Newey (1998), to reduce the sampling variation in (4) by subtracting from it the fitted value associated with its regression onto the infinite-dimensional vector of unconditional moment functions implied by (5)[6]:

$$\begin{aligned} s(Z,\theta_0) &= \frac{D}{p_0(X)}\psi(Z,\beta_0)-\mathbb{E}^*\left[\frac{D}{p_0(X)}\psi(Z,\beta_0)\middle|\frac{D}{p_0(X)}-1;X\right] \\ &= \frac{D}{p_0(X)}\psi(Z,\beta_0)-\frac{q(X;\beta_0)}{p_0(X)}(D-p_0(X)). \end{aligned}$$

[6]The notation $\mathbb{E}^*[Y|X;Z]$ denotes the (mean squared error minimizing) linear predictor of Y given X *within a subpopulation homogenous in* Z:

$$\mathbb{E}^*[Y|X;Z]=X'\pi(Z), \quad \pi(Z)=\mathbb{E}[XX'|Z]^{-1}\times\mathbb{E}[XY|Z].$$

Wooldridge (1999b, Section 4) collected some useful results on conditional linear predictors. See also Newey (1990) and Brown and Newey (1998).

That this population residual is equal to the efficient score function derived by Robins, Rotnitzky, and Zhao (1994) strongly suggests an equivalence between the generalized method-of-moments (GMM) problem defined by restrictions (4) and (5) and the MAR setup outlined above. One way to formally show this is to verify that the efficiency bounds for β_0 in the two problems coincide.[7] The bound for β_0 under the MAR setup is given in (2) above, while under the moment problem, it is established by the following theorem.

THEOREM 2.1—GMM Equivalence: *Suppose that* (i) *the distribution of* Z *has a known, finite support*, (ii) *there is some* $\beta_0 \in \mathcal{B} \subset \mathbb{R}^K$ *and* $\rho_0 = (\rho_1, \dots, \rho_L)'$, *where* $\rho_l = p_0(x_l) \in [\kappa, 1]$ *for each* $l = 1, \dots, L$ *and some* $0 < \kappa < 1$ (*with* $\mathcal{X} = \{x_1, \dots, x_L\}$ *the known support of* X) *such that restrictions* (4) *and* (5) *hold*, (iii) Λ_0 *and* $\mathcal{I}_{\mathrm{m}}(\beta_0) = \Gamma_0'\Lambda_0^{-1}\Gamma_0$ *are nonsingular*, *and* (iv) *other regularity conditions hold* (*cf*. Chamberlain (1992b, *Section* 2)), *then* $\mathcal{I}_{\mathrm{m}}(\beta_0)$ *is the Fisher information bound for* β_0.

All proofs are provided in the Supplemental Material (Graham (2011)).

The proof of Theorem 2.1 involves only some tedious algebra and a straightforward application of Lemma 2 of Chamberlain (1987). Assuming that Z has known, finite support makes the problem fully parametric. The unknown parameters are the probabilities associated with each possible realization of Z, the values of the propensity score at each of the L mass points of the distribution of X, $\rho_0 = (\rho_1, \dots, \rho_L)'$, and the parameter of interest, β_0.

The multinomial assumption is not apparent in the form of $\mathcal{I}_{\mathrm{m}}(\beta_0)$, which involves only conditional expectations of certain functions of the data. This suggests that the bound holds in general, since any F_0 which satisfies (4) and (5) can be arbitrarily well approximated by a multinomial distribution also satisfying the restrictions. Chamberlain (1992a, Theorem 1) demonstrated that this is indeed the case. Therefore, $\mathcal{I}_{\mathrm{m}}(\beta_0)^{-1}$ is the maximal asymptotic precision, in the sense of Hájek's (1972) local minimax approach to efficiency, with which β_0 can be estimated when the only prior restrictions on F_0 are (4) and (5). Since this variance bound coincides with (2), I conclude that (4) and (5) exhaust all of the useful prior restrictions implied by the MAR setup.[8]

The connection between semiparametrically efficient estimation of moment condition models with missing data and augmented systems of moment restrictions has been noted previously for the special case of data missing completely

[7]An alternative approach to showing equivalency would involve verifying Newey's (2004) moment spanning condition for efficiency.

[8]A referee made the insightful observation that the moment condition model (4) and (5) and the MAR setup are equivalent in the stronger sense that they impose identical restrictions on the observed data. This, of course, also implies that they contain identical information on β_0. The complete data vector is given by (D, X, Y_1), with only $(D, X, Y) = (D, X, DY_1)$ observed. Since Y_1 is not observed whenever $D = 0$, we are free to specify its conditional distribution given X and $D = 0$ as desired. Choosing $Y_1|X, D = 0 \overset{D}{\sim} Y_1|X, D = 1$ ensures conditional independence (Assumption 1.3). Manipulating the identifying moment (4), we then have, writing $\psi(Z, \beta_0) =$

at random (MCAR). In that case, Assumptions 1.1–1.4 hold with $p_0(X)$ equal to a (perhaps known) constant. Newey (1994a) showed that an efficient estimate of β_0 can be based on the pair of moment restrictions

$$\mathbb{E}[D\psi(Z,\beta_0)]=0, \quad \mathbb{C}(D,q(X;\beta_0))=0,$$

with $q(X;\beta)$ as defined above. Hirano, Imbens, and Ridder (2003) discussed a related example with X binary and the data also MCAR. In their example, efficient estimation is possible with only a finite number of unconditional moment restrictions. Theorem 2.1 provides a formal generalization of the Newey (1994a) and Hirano, Imbens, and Ridder (2003) examples to the missing at random (MAR) case.

The method-of-moments formulation of the MAR setup provides a useful framework for understanding several apparent paradoxes found in the missing data literature. As a simple example, consider Hahn's (1998, pp. 324–325) result that projection onto a known propensity score may be harmful for estimation of $\beta_0=\mathbb{E}[Y_1]$. Formally, he showed that, for $p_0(x)=Q_0$ constant in x and known, the complete-case estimator, $\widehat{\beta}_{\mathrm{cc}}=\sum_{i=1}^N D_iY_{1i}/\sum_{i=1}^N D_i$, while consistent, is inefficient. Observe that for the constant propensity score case, $\widehat{\beta}_{\mathrm{cc}}$ is the sample analog of the population solution to (4). It consequently makes no use of any information contained in the auxiliary moment (5). However, that moment will be informative for β_0 if $q(x;\beta_0)=\mathbb{E}[Y_1|x]-\beta_0$ varies with x, consistent with Hahn's (1998) finding that the efficiency loss associated with $\widehat{\beta}_{\mathrm{cc}}$ is proportional to $\mathbb{V}(q(X;\beta_0))$. Similar reasoning explains why weighting by the (inverse of) the known propensity score is generally inefficient (cf. Robins, Rotnitzky, and Zhao (1994), Hirano, Imbens, and Ridder (2003), Wooldridge (2007)). The known weights estimator ignores the information contained in (5).

That smoothness and exclusion priors on the propensity score do not lower the variance bound also has a GMM interpretation. Consider the case where the propensity score belongs to a parametric family $p(X;\eta_0)$. If η_0 is known, then an efficient GMM estimator based on (4) and (5) is given by the solution

$\psi(X,Y_1,\beta_0)$,

$$\begin{aligned}\mathbb{E}\left[\frac{D}{p_0(X)}\psi(X,Y,\beta_0)\right]&=\mathbb{E}\left[p_0(X)\mathbb{E}\left[\frac{D}{p_0(X)}\psi(X,DY_1,\beta_0)\middle|X,D=1\right]\right]\\&=\mathbb{E}\big[\mathbb{E}[\psi(X,DY_1,\beta_0)|X,D=1]\big]\\&=\mathbb{E}\big[\mathbb{E}[\psi(X,Y_1,\beta_0)|X]\big],\end{aligned}$$

which yields (1). Finally, the auxiliary restriction (5) ties down the conditional distribution of D given X and ensures Assumption 1.4 is satisfied. I thank Michael Jansson for several helpful discussions on this point.

to

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} s(\eta_0, \widehat{q}, \widehat{\beta}) &= \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{D_i}{p(X_i;\eta_0)}\psi(Z_i,\widehat{\beta}) - \frac{\widehat{q}(X_i;\widehat{\beta})}{p(X_i;\eta_0)}(D_i - p(X_i;\eta_0))\right\} \\ &= 0,\end{aligned}$$

with $\widehat{q}(x;\widehat{\beta})$ a consistent nonparametric estimate of $\mathbb{E}[\psi(Z,\beta_0)|x]$. Now consider the effect of replacing η_0 with the consistent estimate $\widehat{\eta}$. From Newey and McFadden (1994, Theorem 6.2), this replacement does not change the first order asymptotic sampling distribution of $\widehat{\beta}$ because $\mathbb{E}[\partial s(\eta_0, q_0, \beta_0)/\partial\eta'] = 0$. Furthermore, if the known propensity score is replaced by a consistent nonparametric estimate, $\widehat{p}(x)$, then the sampling distribution of $\widehat{\beta}$ is also unaffected (Newey (1994b, Proposition 3, p. 1360)). Since the M-estimate of β_0 based on its efficient score function has the same asymptotic sampling distribution whether the propensity score is set equal to the truth or, instead, to a noisy, but consistent, estimate, knowledge of its form cannot increase the precision with which β_0 can be estimated.

Another intuition for redundancy of knowledge of the propensity score can be found by inspecting the information bound for the multinomial problem. Under the conditions of Theorem 2.1, calculations provided in the Supplemental Material (Graham (2011)) imply that the GMM estimates of β_0 and ρ_0 (recall that ρ_0 contains the values for the propensity score at each of the mass points of the distribution of X) have an asymptotic sampling distribution of

$$\sqrt{N}\left(\begin{bmatrix}\widehat{\rho}\\ \widehat{\beta}\end{bmatrix} - \begin{bmatrix}\rho_0\\ \beta_0\end{bmatrix}\right) \xrightarrow{D} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\mathcal{I}_{\mathrm{m}}(\rho_0)^{-1} & 0\\ 0 & \mathcal{I}_{\mathrm{m}}(\beta_0)^{-1}\end{bmatrix}\right),$$

with $\mathcal{I}_{\mathrm{m}}(\beta_0)$ as defined in (2) and $\mathcal{I}_{\mathrm{m}}(\rho_0)$ as defined in the Supplement Material. As is well known, under block diagonality, sampling error in $\widehat{\rho}$ does not affect, at least to first order, the asymptotic sampling properties of $\widehat{\beta}$. While block diagonality is formally only a feature of the multinomial problem, the result nonetheless provides another useful intuition for understanding why prior knowledge of the propensity score is not valuable asymptotically.

Finally, the redundancy of knowledge of the propensity score combined with the structure of the equivalent GMM problem, suggests why the IPW estimator based on a nonparametric estimate of the propensity score is semiparametrically efficient (Hirano, Imbens, and Ridder (2003)): when a nonparametric estimate of the propensity score is used, the sample analogs of both (4) and (5) are satisfied. In contrast, the IPW estimator based on a parametric estimate of the propensity score will only satisfy a finite number of the moment conditions implied by (5); hence, while it will be more efficient than the estimator that weights by the true propensity score (e.g., Wooldridge (2007)), it will be less efficient than the one proposed by Hirano, Imbens, and Ridder (2003).

3. SEMIPARAMETRIC FUNCTIONAL RESTRICTIONS

Consider the MAR setup augmented by Assumption 1.5. To the best of my knowledge, the maximal asymptotic precision with which β_0 can be estimated in this model has not been previously characterized. To calculate the bound for this problem, I first consider the conditional moment problem defined by (4), (5), and

$$\mathbb{E}\big[\rho(Z, \delta_0, h_0(X_2); \beta_0)|X\big] = 0, \tag{6}$$

with $\rho(Z, \delta_0, h_0(X_2); \beta_0) = \psi(Z, \beta_0) - q(x, \delta_0, h_0(x_2); \beta_0)$. I apply Chamberlain's (1992a) approach to this problem to calculate a variance bound for β_0. I then show that this bound coincides with the semiparametric efficiency bound for the problem defined by restriction (1) and Assumptions 1.1–1.5 using the methods of Bickel, Klaassen, Ritov, and Wellner (1993). The value of first considering the conditional moment problem is that it provides a conjecture for the form of the efficient influence function, therefore sidestepping the need to directly calculate what is evidently a complicated projection.

To present these results, I begin by letting

$$\begin{aligned}
&q_0(X) = q(X, \delta_0, h_0(X_2); \beta_0),\\
&\rho(Z; \beta_0) = \psi(Z, \beta_0) - q_0(X),\\
&\underset{P\times P}{\Upsilon_0^h}(X_2) = \mathbb{E}\left[D\left(\frac{\partial q_0(X)}{\partial h'}\right)' \Sigma_0(X)^{-1}\left(\frac{\partial q_0(X)}{\partial h'}\right)\Big| X_2\right],\\
&\underset{K\times J}{\Upsilon_0^{h\delta}}(X_2) = \mathbb{E}\left[D\left(\frac{\partial q_0(X)}{\partial h'}\right)' \Sigma_0(X)^{-1}\left(\frac{\partial q_0(X)}{\partial \delta'}\right)\Big| X_2\right],\\
&\underset{K\times J}{G_0}(X) = \frac{\partial q_0(X)}{\partial \delta'} - \left(\frac{\partial q_0(X)}{\partial h'}\right)\Upsilon_0^h(X_2)^{-1}\Upsilon_0^{h\delta}(X_2),\\
&\underset{K\times P}{H_0}(X_2) = \mathbb{E}\left[\frac{\partial q_0(X)}{\partial h'}\Big| X_2\right],\\
&\underset{J\times J}{\mathcal{I}_{\mathrm{m}}^{\mathrm{f}}}(\delta_0) = \mathbb{E}[DG_0(X)'\Sigma_0(X)^{-1}G_0(X)],
\end{aligned}$$

and

$$\begin{aligned}
\underset{K\times K}{\Xi_0} = {}& \mathbb{E}[H_0(X_2)\Upsilon_0^h(X_2)^{-1}H_0(X_2)'] + \mathbb{E}[G_0(X)]\mathcal{I}_{\mathrm{m}}^{\mathrm{f}}(\delta_0)^{-1}\mathbb{E}[G_0(X)]'\\
&+ \mathbb{E}[q_0(X)q_0(X)'].
\end{aligned}$$

The variance bound for β_0 in the conditional moment problem defined by (4), (5), and (6) is established by the following theorem.

THEOREM 3.1—Efficiency With Functional Restrictions, Part 1: *Suppose that* (i) *the distribution of* Z *has a known, finite support*, (ii) *there is some* $\beta_0 \in \mathcal{B} \subset \mathbb{R}^K$, $\rho_0 = (\rho_1, \dots, \rho_L)'$, *where* $\rho_l = p_0(x_l) \in [\kappa, 1]$ *for each* $l = 1, \dots, L$ *and some* $0 < \kappa < 1$ (*with* $\mathcal{X} = \{x_1, \dots, x_L\}$ *the known support of* X), $\delta_0 \in \mathcal{D} \subset \mathbb{R}^J$, *and* $h_0(x_{2,m}) = \lambda_{0,m} \in \mathcal{L} \subset \mathbb{R}^P$ *for each* $m = 1, \dots, M$ (*with* $\mathcal{X}_2 = \{x_{2,1}, \dots, x_{2,M}\}$ *the known support of* X_2) *such that restrictions* (4), (5), *and* (6) *hold*, (iii) Ξ_0 *and* $\mathcal{I}^{\mathrm{f}}_{\mathrm{m}}(\beta_0) = \Gamma_0'\Xi_0^{-1}\Gamma_0$ *are nonsingular, and* (iv) *other regularity conditions hold* (*cf.* Chamberlain (1992b, *Section* 2)), *then* $\mathcal{I}^{\mathrm{f}}_{\mathrm{m}}(\beta_0)$ *is the Fisher information bound for* β_0.

Note that if $X_1 = \emptyset$ and $X_2 = X$ such that $\mathbb{E}[\psi(Z, \beta_0)|x]$ is unrestricted, then $\mathcal{I}^{\mathrm{f}}_{\mathrm{m}}(\beta_0)$ simplifies to $\mathcal{I}_{\mathrm{m}}(\beta_0)$ above. Therefore, Theorem 2.1 may be viewed as a special case of Theorem 3.1. As with Theorem 2.1, the validity of the bound for the non-multinomial case follows from Theorem 1 of Chamberlain (1992a).

The form of Ξ_0 suggests a candidate efficient influence function of

$$(7) \qquad \phi^{\mathrm{f}}_{\beta}(Z, \eta_0, \beta_0) = \Gamma_0^{-1}\Biggl\{DH_0(X_2)Y_0^h(X_2)^{-1}\left(\frac{\partial q_0(X)}{\partial h'}\right)' \times \Sigma_0(X)^{-1}\rho(Z;\beta_0) + D\mathbb{E}[G_0(X)]\mathcal{I}^{\mathrm{f}}_{\mathrm{m}}(\delta_0)^{-1}G_0(X)' \times \Sigma_0(X)^{-1}\rho(Z;\beta_0) + q(X;\beta_0)\Biggr\},$$

where $\eta = (h, \delta, H, Y^h, Y^{h\delta}, \Sigma, \overline{G})$, with $\overline{G} = \mathbb{E}[G(X)]$. Note that each of the three components of (7) is mutually uncorrelated. The next theorem verifies that (7) is the efficient influence function under the MAR setup with Assumption 1.5 also imposed.

THEOREM 3.2—Efficiency With Functional Restrictions, Part 2: *The semiparametric efficiency bound for* β_0 *in the problem defined by restriction* (1) *and Assumptions* 1.1–1.5 *is equal to* $\mathcal{I}^{\mathrm{f}}_{\mathrm{m}}(\beta_0)$ *with an efficient influence function of* $\phi^{\mathrm{f}}_{\beta}(Z, \eta_0, \beta_0)$.

Theorem 3.1 implies that restriction (6) can be exploited to more efficiently estimate β_0. However, its use also carries risk: if false, yet nevertheless erroneously maintained by the data analyst, an inconsistent estimate of β_0 may result. This tension, between efficiency and robustness is formalized by the next two propositions, which together provide guidance as to whether prior information of the type given by Assumption 1.5 should be utilized in practice.

The first proposition characterizes the magnitude of the efficiency gain associated with correctly exploiting Assumption 1.5. Define

$$\underset{K\times 1}{\xi_1(Z, \eta_0, \beta_0)} = D\Bigg\{\frac{I_K}{p_0(X)} - H_0(X_2)Y_0^h(X_2)^{-1}\left(\frac{\partial q_0(X)}{\partial h'}\right)'\Sigma_0(X)^{-1}\Bigg\}\rho(Z;\beta_0),$$

$$\underset{J\times 1}{\xi_2(Z, \eta_0, \beta_0)} = DG_0(X)'\Sigma_0(X)^{-1}\rho(Z;\beta_0).$$

PROPOSITION 3.1: *Under* (1) *and Assumptions* 1.1–1.5,

$$\mathcal{I}_{\mathrm{m}}(\beta_0)^{-1} - \mathcal{I}_{\mathrm{m}}^{\mathrm{f}}(\beta_0)^{-1} = \Gamma_0^{-1}(\mathbb{V}(\xi_1) - \mathbb{C}(\xi_1, \xi_2')\mathbb{V}(\xi_2)^{-1}\mathbb{C}(\xi_1, \xi_2')')\Gamma_0^{-1\prime} \geq 0. \tag{8}$$

Equation (8) has an intuitive interpretation. The first term in parentheses,

$$\mathbb{V}(\xi_1) = \mathbb{E}\left[\frac{\Sigma_0(X)}{p_0(X)} - H_0(X_2)Y_0^h(X_2)^{-1}H_0(X_2)'\right],$$

equals the asymptotic variance reduction that would be available by additionally imposing restriction (6) if δ_0 were *known*.

The additional (asymptotic) sampling uncertainty induced by having to estimate δ_0 is captured by the second term

$$\mathbb{C}(\xi_1, \xi_2)\mathbb{V}(\xi_2)^{-1}\mathbb{C}(\xi_1, \xi_2) = \mathbb{E}[G_0(X)]\mathcal{I}_{\mathrm{m}}^{\mathrm{f}}(\delta_0)^{-1}\mathbb{E}[G_0(X)]',$$

where $\mathcal{I}_{\mathrm{m}}^{\mathrm{f}}(\delta_0)$ is the information bound for δ_0 in the semiparametric regression problem (cf. Chamberlain (1992a)):

$$D\psi(Z, \beta_0) = Dq(X, \delta_0, h_0(X_2); \beta_0) + DV,$$
$$\mathbb{E}[V|X, D=1] = \mathbb{E}[V|X] = 0.$$

The more precisely determined is δ_0, the greater the efficiency gain from imposing Assumption 1.5. The size of $\mathbb{E}[G_0(X)]$ also governs the magnitude of the efficiency gain. Conditional on X_2, $(\frac{\partial q_0(X)}{\partial h'})Y_0^h(X_2)^{-1}Y_0^{h\delta}(X_2)$ is a weighted linear predictor of $\frac{\partial q_0(X)}{\partial \delta'}$ given $\frac{\partial q_0(X)}{\partial h'}$ *in the* $D=1$ *subpopulation*. That is,[9]

$$\left(\frac{\partial q_0(X)}{\partial h'}\right) \Upsilon_0^h(X_2)^{-1}\Upsilon_0^{h\delta}(X_2) = \mathbb{E}^*_{\Sigma_0(X)}\left[\frac{\partial q_0(X)}{\partial \delta'}\middle|\frac{\partial q_0(X)}{\partial h'}; X_2, D=1\right],$$

and hence $G_0(X)$ is equal to the difference between $\frac{\partial q_0(X)}{\partial \delta'}$ and its predicted value based on a weighted least squares regression in the $D=1$ subpopulation. The average of these differences, $\mathbb{E}[G_0(X)]$, is taken across the *entire* population; it will be large in absolute value when the distribution of X_1 conditional on X_2 differs in the $D=1$ versus $D=0$ subpopulations. This will occur whenever X_1 is highly predictive for missingness (conditional on X_2). In such situations, the efficiency costs of sampling uncertainty in $\widehat{\delta}$ are greater (relative to the known δ_0 case) because estimation of β_0 requires greater levels of extrapolation.

An example clarifies the discussion given above. Assume that $\psi(Z,\beta_0) = Y_1 - \beta_0$ with

$$q(X,\delta_0,h_0(X_2);\beta_0) = X_1'\delta_0 + h_0(X_2) - \beta_0.$$

This is the model considered by Wang, Linton, and Härdle (2004). In addition to being of importance in its own right, it provides insight into the program evaluation problem (where the means of two missing outcomes, as opposed to just one, need to be estimated). The Wang, Linton, and Härdle (2004) prior restriction includes the condition that $\mathbb{V}(Y_1|X) = \sigma_1^2$ is constant in X. For clarity of exposition, I also assume homoscedasticity holds, but that this fact is not known by the econometrician. Let $e_0(X_2) = \mathbb{E}[p(X)|X_2] = \Pr(D=1|X_2)$; specializing the general results given above to this model and evaluating (8) gives

$$\begin{aligned}\mathcal{I}_{\mathrm{m}}(\beta_0)^{-1} - \mathcal{I}^{\mathrm{f}}_{\mathrm{m}}(\beta_0)^{-1} = \sigma_1^2\Bigg\{&\mathbb{E}\left[\mathbb{E}\left[\frac{1}{p(X)}\middle|X_2\right] - \frac{1}{e_0(X_2)}\right] \\ &- \big(\mathbb{E}\big[\mathbb{E}[X_1|X_2] - \mathbb{E}[X_1|X_2, D=1]\big]\big)' \\ &\times \big(\mathbb{E}\big[\mathbb{E}[X_1|X_2] - \mathbb{E}[X_1|X_2, D=1]\big]\big) \\ &/\mathbb{E}[e_0(X_2)\mathbb{V}(X_1|X_2, D=1)]\Bigg\} \geq 0,\end{aligned}$$

[9]The notation $\mathbb{E}^*_{\omega(X)}[Y|X; Z, D=1]$ denotes the weighted conditional linear predictor

$$\mathbb{E}^*_{\omega(X)}[Y|X; Z, D=1] = X\mathbb{E}[DX\omega(X)^{-1}X'|Z]^{-1} \times \mathbb{E}[DX\omega(X)^{-1}Y|Z].$$

This is the population analog of the fitted value from a generalized least squares regression in a subpopulation homogenous in Z and with $D=1$.

which shows that the efficiency gain associated with correctly exploiting Assumption 1.5 reflects three forces. First, substantial convexity in $p(X)^{-1}$, which will occur when overlap is limited, increases the efficiency gain.[10] This gain reflects the semiparametric restriction allowing for extrapolation in the presence of conditional covariate imbalance. The next two effects reflect the fact that the first source of efficiency gain is partially nullified by having to estimate δ_0. If X_1 varies strongly given X_2 in the $D=1$ subpopulation, then the information for δ_0 is large, which, in turn, increases the precision with which β_0 may be estimated. On the other hand, if there are large (average) differences in the conditional mean of X_1 given X_2 across the $D=1$ and $D=0$ subpopulations, then estimating β_0 requires greater extrapolation, which, when δ_0 is unknown, decreases the precision with which it may be estimated.

Proposition 3.1 provides insight into when correctly imposing Assumption 1.5 is likely to be informative. A related question concerns the consequences of misspecifying the form of $q(X, \delta, h(X_2); \beta)$. Under such misspecification, the conditional moment restriction (6) will be invalid. Nevertheless, the efficient score function may continue to have an expectation of zero at $\beta = \beta_0$. This suggest that an M-estimator based on an estimate of the efficient score function may be consistent even if Assumption 1.5 does not hold. The following proposition provides one set of conditions under which such a robustness property holds.

PROPOSITION 3.2—Double Robustness: *Let* $q_*(X) = q(X, \delta_*, h_*(X_2); \beta_0)$ *with* δ_* *and* $h_*(X_2)$ *arbitrary, let* $\rho_*(Z; \beta_0) = \psi(Z, \beta_0) - q_*(X)$, *and redefine* $\Sigma_0(X) = \mathbb{V}(\rho_*(Z; \beta_0)|X)$, $H_0(X_2) = \mathbb{E}[\frac{\partial q_*(X)}{\partial h'}|X_2]$ *and* $\Upsilon_0^h(X_2)$, $\Upsilon_0^{h\delta}(X_2)$, *and* $\overline{G}_0$ *similarly. Under restriction* (1) *and Assumptions* 1.1–1.4, $\phi_\beta^{\mathrm{f}}(Z, \eta, \beta_0)$ *is mean zero if either* (i) $\beta = \beta_0$, $\eta = \eta_0$ *and Assumption* 1.5 *holds or* (ii) $\beta = \beta_0$, $\eta = \eta_* = (h_*, \delta_*, H_0, \Upsilon_0^h, \Upsilon_0^{h\delta}, \Sigma_0, \overline{G}_0)$, *and* (a) $p_0(x) = e_0(x_2)$ *for all* $x \in \mathcal{X}$, (b) $\Sigma_0(x) = \Theta_0(x_2)$ *for all* $x \in \mathcal{X}$, *and* (c) *at least one element of* $h_*(x_2)$ *enters linearly in each row of* $q_*(X)$.

Note that there is a tension between the robustness property of Proposition 3.2 and the efficiency gain associated with Assumption 1.5. Mean-zeroness of $\phi_\beta^{\mathrm{f}}(Z, \eta, \beta_0)$ under misspecification requires that those variables entering $q(X, \delta, h(X_2); \beta_0)$ parametrically do not affect either the probability of missingness or the conditional variance of the moment function (1). Under such conditions, an estimator based on $\phi_\beta^{\mathrm{f}}(Z, \eta, \beta_0)$ will perform no better, at least asymptotically, than one based on the efficient score function derived by Robins, Rotnitzky, and Zhao (1994). In particular, we have the following implication.

[10]When some subpopulations have low propensity scores, $\mathbb{E}[1/p(X)|X_2] - 1/\mathbb{E}[p(X)|X_2]$ will tend to be large (Jensen's inequality).

COROLLARY 3.1: *Under the conditions of part* (ii) *of Proposition* 3.2,

$$\mathcal{I}_{\mathrm{m}}(\beta_0)^{-1} - \mathcal{I}_{\mathrm{m}}^{\mathrm{f}}(\beta_0)^{-1} = 0.$$

Collectively Propositions 3.1 and 3.2 suggest that estimation while maintaining Assumption 1.5 will be most valuable when the econometrician is highly confident in the imposed semiparametric restriction.

REFERENCES

ANGRIST, J. D., AND A. B. KRUEGER (1992): "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables With Moments From Two Samples," *Journal of the American Statistical Association*, 87, 328–336. [438]

BANG, H., AND J. M. ROBINS (2005): "Doubly Robust Estimation in Missing Data and Causal Inference Models," *Biometrics*, 61, 962–972. [438]

BICKEL, P. J., C. A. J. KLAASSEN, Y. RITOV, AND J. A. WELLNER (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. New York: Springer-Verlag. [445]

BROWN, B. W., AND W. K. NEWEY (1998): "Efficient Semiparametric Estimation of Expectations," *Econometrica*, 66, 453–464. [441]

CHAMBERLAIN, G. (1987): "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334. [437,442]

——— (1992a): "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60, 567–596. [442,445-447]

——— (1992b): "Comment: Sequential Moment Restrictions in Panel Data," *Journal of Business & Economic Statistics*, 10, 20–26. [442,446]

CHEN, X., H. HONG, AND E. T. TAMER (2005): "Measurement Error Models With Auxiliary Data," *Review of Economic Studies*, 72, 343–366. [438]

CHEN, X., H. HONG, AND A. TAROZZI (2004): "Semiparametric Efficiency in GMM Models of Nonclassical Measurement Errors, Missing Data And Treatment Effects," Discussion Paper 1644, Cowles Foundation. [439]

——— (2008): "Semiparametric Efficiency in GMM Models With Auxiliary Data," *The Annals of Statistics*, 36, 808–843. [438]

CHENG, P. E. (1994): "Nonparametric Estimation of Mean Functionals With Data Missing at Random," *Journal of the American Statistical Association*, 89, 81–87. [438,440]

ENGLE, R. F., C. W. J. GRANGER, J. RICE, AND A. WEISS (1986): "Semiparametric Estimates of the Relation Between Weather and Electricity Sales," *Journal of the American Statistical Association*, 81, 310–320. [440]

GRAHAM, B. S. (2011): "Supplement to 'Efficiency Bounds for Missing Data Models With Semiparametric Restrictions'," *Econometrica Supplemental Material*, 79, http://www.econometricsociety.org/ecta/Supmat/7379_Proofs.pdf. [442,444]

GRAHAM, B. S., C. PINTO, AND D. EGEL (2010): "Inverse Probability Tilting and Missing Data Problems," Working Paper 13981, NBER. [439]

HAHN, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66, 315–331. [438-440,443]

——— (2004): "Functional Restriction and Efficiency in Causal Inference," *Review of Economics and Statistics*, 86, 73–76. [439]

HÁJEK, J. (1972): "Local Asymptotic Minimax and Admissibility in Estimation," in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, ed. by L. M. Le Cam, J. Neyman, and E. L. Scott. Berkeley: University of California Press, 175–194. [442]

HIRANO, K., G. W. IMBENS, AND G. RIDDER (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189. [438-441, 443-445]

HITOMO, K., Y. NISHIYAMA, AND R. OKUI (2008): "A Puzzling Phenomenon in Semiparametric Estimation Problems With Infinite-Dimensional Nuisance Parameters," *Econometric Theory*, 24, 1717–1728. [439]

ICHIMURA, H., AND O. LINTON (2005): "Asymptotic Expansions for Some Semiparametric Program Evaluation Estimators," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. W. K. Andrews and J. H. Stock. Cambridge: Cambridge University Press, 149–170. [440]

IMBENS, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *Review of Economics and Statistics*, 86, 4–29. [438,440]

IMBENS, G. W., W. K. NEWEY, AND G. RIDDER (2005): "Mean-Square-Error Calculations for Average Treatment Effects," Working Paper 05.34, IEPR. [438]

LITTLE, R. J. A., AND D. B. RUBIN (2002): *Statistical Analysis With Missing Data*. Hoboken, NJ: Wiley. [440]

NEWEY, W. K. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99–135. [441]

——— (1994a): "Series Estimation of Regression Functionals," *Econometric Theory*, 10, 1–28. [441,443]

——— (1994b): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382. [444]

——— (2004): "Efficient Semiparametric Estimation via Moment Restrictions," *Econometrica*, 72, 1877–1897. [442]

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. 4, ed. by R. F. Engle and D. L. McFadden. Amsterdam: North-Holland, 2111–2245. [444]

PROKHOROV, A., AND P. J. SCHMIDT (2009): "GMM Redundancy Results for General Missing Data Problems," *Journal of Econometrics*, 151, 47–55. [439]

ROBINS, J. M., AND A. ROTNITZKY (1995): "Semiparametric Efficiency in Multivariate Regression Models," *Journal of the American Statistical Association*, 90, 122–129. [438,439]

ROBINS, J. M., F. HSIEH, AND W. NEWEY (1995): "Semiparametric Efficient Estimation of a Conditional Density Function With Missing or Mismeasured Covariates," *Journal of the Royal Statistical Society, Ser. B*, 57, 409–424. [438,439]

ROBINS, J. M., A. ROTNITZKY, AND L. P. ZHAO (1994): "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed," *Journal of the American Statistical Association*, 89, 846–866. [438-443,449]

——— (1995): "Analysis of Semiparametric Regression-Models for Repeated Outcomes in the Presence of Missing Data," *Journal of the American Statistical Association*, 90, 106–121. [438]

SCHARFSTEIN, D. O., A. ROTNITZKY, AND J. M. ROBINS (1999): "Rejoinder," *Journal of the American Statistical Association*, 94, 1135–1146. [438]

TSIATIS, A. A. (2006): *Semiparametric Theory and Missing Data*. New York: Springer. [438]

WANG, Q., O. LINTON, AND W. HÄRDLE (2004): "Semiparametric Regression Analysis With Missing Response at Random," *Journal of the American Statistical Association*, 99, 334–345. [440,448]

WOOLDRIDGE, J. M. (1999a): "Asymptotic Properties of Weighted M-Estimators for Variable Probability Samples," *Econometrica*, 67, 1385–1406. [438]

——— (1999b): "Distribution-Free Estimation of Some Nonlinear Panel Data Models," *Journal of Econometrics*, 90, 77–97. [441]

——— (2002): "Inverse Probability Weighted M-Estimators for Sample Selection, Attrition and Stratification," *Portuguese Economic Journal*, 1, 117–139. [438]

——— (2007): "Inverse Probability Weighted Estimation for General Missing Data Problems," *Journal of Econometrics*, 141, 1281–1301. [439,441,443,445]

Dept. of Economics, New York University, 19 West 4th Street 6FL, New York, NY 10012, U.S.A. and National Bureau of Economic Research; bryan.graham@nyu.edu.

Manuscript received August, 2007; final revision received June, 2010.