The Annals of Statistics
2012, Vol. 40, No. 6, 3031–3049
DOI: 10.1214/12-AOS1061

THE TRANSFER PRINCIPLE: A TOOL FOR COMPLETE CASE ANALYSIS

BY HIRA L. KOUL, URSULA U. MÜLLER[1] AND ANTON SCHICK[2]

Michigan State University, Texas A&M University and Binghamton University

This paper gives a general method for deriving limiting distributions of complete case statistics for missing data models from corresponding results for the model where all data are observed. This provides a convenient tool for obtaining the asymptotic behavior of complete case versions of established full data methods without lengthy proofs.

The methodology is illustrated by analyzing three inference procedures for partially linear regression models with responses missing at random. We first show that complete case versions of asymptotically efficient estimators of the slope parameter for the full model are efficient, thereby solving the problem of constructing efficient estimators of the slope parameter for this model. Second, we derive an asymptotically distribution free test for fitting a normal distribution to the errors. Finally, we obtain an asymptotically distribution free test for linearity, that is, for testing that the nonparametric component of these models is a constant. This test is new both when data are fully observed and when data are missing at random.

1. Introduction.

The basis for regression is a response variable Y and a covariate vector X which are linked via the formula $Y = r(X) + \varepsilon$, where r is a regression function and ε is an error variable. The analysis is then carried out based on independent copies $(X_1, Y_1), \dots, (X_n, Y_n)$ of the pair (X, Y). We refer to this as the *full model*. In applications, however, responses may be missing. The base observation is then a triple $(X, \delta Y, \delta)$, where δ is an indicator variable with $E[\delta] = P(\delta = 1) > 0$. The interpretation is that for $\delta = 1$, one observes the pair (X, Y), while for $\delta = 0$, one only observes the covariate X. The analysis is now based on independent copies $(X_1, \delta_1 Y_1, \delta_1), \dots, (X_n, \delta_n Y_n, \delta_n)$ of the observation $(X, \delta Y, \delta)$. An accepted way of analyzing such data is by imputing the missing responses. Here we take a closer look at *complete case analysis*. This method ignores the incomplete observations and works with only the $N = \sum_{j=1}^n \delta_j$ completely observed pairs $(X_{i_1}, Y_{i_1}), \dots, (X_{i_N}, Y_{i_N})$. Formally, to each statistic

$$T_n = t_n(X_1, Y_1, \dots, X_n, Y_n)$$

Received August 2011; revised June 2012.

[1]Supported by NSF Grant DMS-09-07014.

[2]Supported by NSF Grant DMS-09-06551.

MSC2010 subject classifications. Primary 62E20; secondary 62G05, 62G10.

Key words and phrases. Transfer principle, missing at random, partially linear models, efficient estimation, martingale transform test for normal errors, testing for linearity.

for the full model there corresponds the *complete case statistic*

$$T_c = t_N(X_{i_1}, Y_{i_1}, \dots, X_{i_N}, Y_{i_N}),$$

which mimics the statistic T_n by treating $(X_{i_1}, Y_{i_1}), \dots, (X_{i_N}, Y_{i_N})$ as if it were a sample of size N from the original setting without missing data.

Our main result gives a simple and useful method for obtaining the asymptotic distribution of T_c. We show that the limiting distribution of T_c coincides with that of $\tilde{T}_n = t_n(\tilde{X}_1, \tilde{Y}_1, \dots, \tilde{X}_n, \tilde{Y}_n)$ where $(\tilde{X}_1, \tilde{Y}_1), \dots, (\tilde{X}_n, \tilde{Y}_n)$ form a random sample drawn from the conditional distribution of (X, Y), given $\delta = 1$; see Remark 2.4. This can be used as follows. One typically knows the limiting distribution $\mathfrak{L}(Q)$ of T_n under each joint distribution Q of X and Y belonging to some model. If the distribution $\tilde{Q}$ of $(\tilde{X}, \tilde{Y})$ belongs to this model, then the limiting distribution of the complete case statistic is $\mathfrak{L}(\tilde{Q})$. We refer to this as the *transfer principle*. It provides a convenient tool for obtaining the asymptotic behavior of complete case versions of established full data methods without (reproducing) lengthy proofs.

Of special interest are statistics T_n that are *asymptotically linear* for a functional T from a class $\mathcal{Q}$ of distributions into $\mathbb{R}^m$ in the sense that if X and Y have joint distribution Q belonging to the model $\mathcal{Q}$, then the expansion

$$T_n = T(Q) + \frac{1}{n}\sum_{j=1}^{n} \psi_Q(X_j, Y_j) + o_P(n^{-1/2})$$

holds. Here ψ_Q is a measurable function into $\mathbb{R}^m$ such that $E[\psi_Q(X, Y)] = 0$ and $E[\|\psi_Q(X, Y)\|^2]$ is finite when X and Y have joint distribution Q. Here and below $\|\cdot\|$ denotes the Euclidean norm. The function ψ_Q is commonly called an *influence function*. From the above expansion we obtain that $n^{1/2}(T_n - T(Q))$ is asymptotically normal with the zero vector as mean and with dispersion matrix $\Sigma(Q) = E[\psi_Q(X, Y)\psi_Q^\top(X, Y)]$. If $\tilde{Q}$ belongs to the model $\mathcal{Q}$, then we have the expansion

$$\tilde{T}_n = T(\tilde{Q}) + \frac{1}{n}\sum_{j=1}^{n} \psi_{\tilde{Q}}(\tilde{X}_j, \tilde{Y}_j) + o_P(n^{-1/2}),$$

and obtain from our main result that

$$T_c = T(\tilde{Q}) + \frac{1}{N}\sum_{j=1}^{n} \delta_j \psi_{\tilde{Q}}(X_j, Y_j) + o_P(n^{-1/2}),$$

see Remark 2.5. From this we immediately derive the expansion

$$T_c = T(\tilde{Q}) + \frac{1}{nE[\delta]}\sum_{j=1}^{n} \delta_j \psi_{\tilde{Q}}(X_j, Y_j) + o_P(n^{-1/2}).$$

Thus, if $\tilde{Q}$ belongs to the model $\mathcal{Q}$ and $T(\tilde{Q})$ equals $T(Q)$, then T_c is asymptotically linear in the model with missing data with influence function $\tilde{\psi}$, where

$$\tilde{\psi}(X, \delta Y, \delta) = \frac{\delta}{E[\delta]} \psi_{\tilde{Q}}(X, Y).$$

We refer to this as the *transfer principle for asymptotically linear statistics*. It yields that $n^{1/2}(T_c - T(Q))$ is asymptotically normal with the zero vector as mean and with dispersion matrix $(1/E[\delta])\Sigma(\tilde{Q})$.

The key to a successful application of the transfer principle is the condition $T(\tilde{Q}) = T(Q)$. Under this condition, $n^{1/2}$-consistency carries over to the complete case statistic. If this condition is not met, the complete case statistic will be biased for estimating $T(Q)$.

For our illustration of the transfer principle we consider the important case where the response Y is *missing at random* (MAR). This means that the indicator δ is conditionally independent of Y, given X, that is,

$$P(\delta = 1|X, Y) = P(\delta = 1|X) = \pi(X) \qquad \text{a.s.}$$

This is a common assumption and reasonable in many applications [see Little and Rubin (2002), Chapter 1]. This model is referred to as the *MAR model*.

It is well known that the complete case analysis does not always perform well and that an approach which imputes missing values often has better statistical properties. See, for example, Chapter 3 of Little and Rubin (2002) for examples where using the complete case approach results in bias or a loss of precision, due to the loss of information. For a discussion of various imputing methods we again refer to Little and Rubin (2002), and also to Müller, Schick and Wefelmeyer (2006), who propose efficient estimators for various regression settings which impute missing and non-missing responses.

Although complete case analysis can lead to the above-mentioned problems, there are situations where it provides useful and optimal inference procedures. Efromovich (2011), for example, considers nonparametric regression with responses missing at random. He shows that his complete case estimator of the regression function is optimal in the sense that it satisfies an asymptotic sharp minimax property. Müller (2009) demonstrates efficiency of a complete case estimator for the parameter vector in the nonlinear regression model.

For simplicity and clarity, we illustrate the above transfer principle using the partially linear regression model. In this model the response Y is linked to covariates U and V via the relation

$$Y = \vartheta^\top U + \rho(V) + \varepsilon, \tag{1.1}$$

with ϑ an unknown m-dimensional vector and ρ an unknown twice continuously differentiable function. The error ε is assumed to have mean zero, finite variance σ^2 and a density f, and is independent of the covariates (U, V), where the random vector U has dimension m and the random variable V takes values in the compact interval [0, 1]. Throughout this paper, we impose the following conditions on the joint distribution G of U and V:

(G1) The covariate V has a density that is bounded and bounded away from zero on [0, 1].

(G2) The covariate vector U satisfies $E[\|U\|^2] < \infty$ and the matrix

$$W_G = E\big[\big(U - \mu_G(V)\big)\big(U - \mu_G(V)\big)^\top\big]$$

is positive definite, where $\mu_G(V) = E[U|V]$.

The requirement involving W_G is needed to identify the parameter ϑ.

One important problem is the *efficient* estimation of the regression parameter ϑ in (1.1). This is addressed in our first illustration of the transfer principle below. The crucial condition for a successful application of the transfer principle, $T(\tilde{Q}) = T(Q)$, is satisfied in this case and, more generally, also for functionals of the triple (ϑ, ρ, f). The MAR assumption and the independence of ε and (U, V) imply that ε and (U, V, δ) are independent. Hence, the regression parameters ϑ and ρ and the error density f stay the same when conditioning on $\delta = 1$. Only the covariate distribution G changes to $\tilde{G}$, the conditional distribution of (U, V) given $\delta = 1$. This argument suggests that inference about the triple (ϑ, ρ, f) should be carried out using a complete case analysis, because the complete case observations are *sufficient* for $(\vartheta, \rho, f, \tilde{G})$ since they carry all the information about these parameters. The covariate pair (U, V) alone, on the other hand, has no information on (ϑ, ρ, f), and hence has no bearing on the inference about these parameters when the response Y is missing at random. The same reasoning also applies to general semiparametric regression models: inference about the regression function and the error distribution should be based on the complete cases only.

In order to obtain an efficient estimator for ϑ we must assume that the error density f has finite Fisher information for location. This means that f is absolutely continuous with a.e. derivative f' such that $J_f = \int \ell_f^2(x) f(x)\,dx$ is finite, where $\ell_f = -f'/f$ is the score function for location. Efficient estimators of ϑ in the full model are characterized by the stochastic expansion

$$\hat{\vartheta}_n = \vartheta + \frac{1}{n}\sum_{j=1}^{n}(J_f W_G)^{-1}\big(U_j - \mu_G(V_j)\big)\ell_f(\varepsilon_j) + o_P\big(n^{-1/2}\big);$$

see, for example, Schick (1993). Because of the structure of the MAR model introduced above, the transfer principle for asymptotically linear statistics yields that the complete case version $\hat{\vartheta}_c$ of an efficient estimator satisfies the expansion

$$\hat{\vartheta}_c = \vartheta + \frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j}{E[\delta]}(J_f W_{\tilde{G}})^{-1}\big(U_j - \mu_{\tilde{G}}(V_j)\big)\ell_f(\varepsilon_j) + o_P\big(n^{-1/2}\big). \tag{1.2}$$

This of course requires that $\tilde{G}$ satisfies the properties (G1) and (G2). This is the case when π is bounded away from zero; see Remark 3.1. Here $\pi(X) = \pi(U, V) = P(\delta = 1|U, V)$.

Although several estimators exist which are efficient in the full partially linear model, to our knowledge no efficient estimators have so far been constructed for the corresponding MAR model. We show in Section 3 that the expansion (1.2) of $\hat{\vartheta}_c$ characterizes asymptotically *efficient* estimators of ϑ in the MAR model. This means that complete case versions of efficient estimators in the full model remain efficient in the MAR model (under appropriate conditions). This result in turn solves the important problem of constructing efficient estimators for ϑ in the partially linear MAR model. For constructions of efficient estimators in the full model (1.1), we refer the reader to Cuzick (1992), Schick (1993, 1996), Bhattacharya and Zhao (1997) and Forrester et al. (2003). Some of these constructions require smoothness assumptions on μ_G. Then the validity of (1.2) requires the same smoothness assumptions on $\mu_{\tilde{G}}$.

The above method of constructing efficient estimators for the finite-dimensional parameter also yields efficient estimators in other semiparametric regression MAR models. The influence function of the complete case version of an estimator efficient for the full model is given by the transfer principle for asymptotically linear estimators. One then only needs to show that this influence function is the efficient influence function for the MAR model. The latter can be done by mimicking the results in Section 3. There we sketch this approach for the partially linear model with additive ρ and for a single index model. Müller (2009) has calculated the efficient influence function for the regression parameter in a nonlinear regression model. Using the transfer principle, one sees that the efficient influence function equals the influence function of the complete case version of an efficient estimator for the full model. This provides a simple derivation of efficient estimators in her model.

We believe that the above *efficiency transfer* is valid for the estimation of other characteristics in the MAR model (1.1). We expect that the efficiency transfer generalizes to the estimation of (smooth) functionals of the triple (ϑ, ρ, f). This includes as important special cases the estimation of the error distribution function, the error variance and other characteristics of f such as quantiles and moments of f. However, further research is needed to crystallize the issues involved.

Next we illustrate the transfer principle on goodness-of-fit and lack-of-fit tests. There is a vast literature on goodness-of-fit tests for fitting an error distribution and lack-of-fit tests for fitting a regression function in fully observable regression models. See, for example, Hart (1997) and the review article by Koul (2006), and the references therein. Here we shall discuss two important examples for the MAR regression models. One pertains to fitting a parametric distribution to the error distribution in (1.1) and the other to testing whether ρ in the model (1.1) is a constant or not. In both examples the proposed tests are complete case analogs of full model tests that are *asymptotically distribution free*, that is, the limiting distribution of the test statistic under the null hypothesis is the same for all members of the null model being fitted. Due to the transfer principle, the same conclusion continues to hold for the proposed tests for the MAR model (1.1).

First, consider the goodness-of-fit testing problem in the model (1.1) and the null hypothesis $H_0 : \varepsilon \sim N(0, \sigma^2)$, for some unknown $0 < \sigma^2 < \infty$. For the full model a residual-based test of this hypothesis was introduced by Müller, Schick and Wefelmeyer (2012) (MSW) adapting a martingale transform test of Khmaladze and Koul (2009) for fitting a parametric family of error distributions in nonparametric regression. In (1.1), the residuals are of the form $\hat{\varepsilon}_j = Y_j - \hat{\vartheta}^\top U_j - \hat{\rho}(V_j)$, where $\hat{\vartheta}$ is a $\sqrt{n}$-consistent estimator of ϑ and $\hat{\rho}$ is a nonparametric estimator of ρ, such as a local smoother based on the covariates V_j and the modified responses $Y_j - \hat{\vartheta}^\top U_j$, or a series estimator. Let $\hat{\sigma} = (\sum_{j=1}^n \hat{\varepsilon}_j^2/n)^{1/2}$ denote the estimator of the standard deviation σ and $\hat{Z}_j = \hat{\varepsilon}_j/\hat{\sigma}$, $j = 1, \dots, n$, denote the standardized residuals. The test statistic of MSW is then

$$T_n = \sup_{t \in \mathbb{R}} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^n \{\mathbf{1}[\hat{Z}_j \le t] - H(\hat{Z}_j \wedge t) h(\hat{Z}_j)\} \right|$$

for some known functions h and H related to the standard normal distribution function and its derivatives; see Section 4, equation (4.3). Here we work with a series estimator of ρ, which is discussed in Section 4 of MSW. This requires no additional assumptions. The test based on T_n is asymptotically distribution free, because under the null hypothesis, T_n converges in distribution to

$$\zeta = \sup_{0 \le t \le 1} |B(t)|, \tag{1.3}$$

where B is a standard Brownian motion. Due to the transfer principle, the complete case version T_c of the above T_n has the same limiting distribution under the null hypothesis. Hence, the null hypothesis is rejected if T_c exceeds the upper α quantile of the distribution of ζ. See Section 4, equation (4.4), and the discussion around it for a detailed description of the complete case variant T_c of the above T_n. From the discussion on optimality of this test in Khmaladze and Koul (2009) and the transfer principle, it follows that the test based on T_c will generally be more powerful than the complete case test based on the Kolmogorov–Smirnov statistic.

Finally, we consider testing whether ρ is constant within the partially linear model, that is, we suppose that the partially linear model (1.1) holds true and test whether the regression function is in fact linear. Here we adapt an approach by Stute, Xu and Zhu (2008) for testing a general parametric model in nonparametric regression, which is based on a weighted residual-based empirical process. For the full model this suggests a test statistic of the form

$$\sup_{t \in \mathbb{R}} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^n W_j 1[\hat{\varepsilon}_{j0} \le t] \right|,$$

where $\hat{\varepsilon}_{j0}$ are the residuals under the null hypothesis obtained by regressing the responses Y_j on the covariates U_j including an intercept, and where W_j are normalized versions of the residuals obtained from regressing $\chi(V_j)$ on the covariates U_j including an intercept, for a suitably chosen function χ. The asymptotic null distribution of this test is that of

$$\zeta_0 = \sup_{0 \le t \le 1} \left| B_0(t) \right|,$$

where B_0 denotes a standard Brownian bridge. This is the first test for this problem in the case of fully observed data. The transfer principle immediately shows that the complete case variant of this test described at (5.1) has the same limiting distribution.

The literature on lack-of-fit testing in the regression model when responses are missing at random is scant. Sun and Wang (2009) establish asymptotic distributional properties of some tests based on marked residual empirical processes for fitting a parametric model to the regression function when data are imputed using the inverse probability method. Sun, Wang and Dai (2009) derive tests to check the hypothesis that the partially linear model (1.1) is appropriate, based on data which are "completed" by imputing estimators for the responses. These tests are compared with tests that ignore the missing data pairs. González-Manteiga and Pérez-González (2006) use imputation to complete the data. They derive tests about linearity of the regression function in a general nonparametric regression model. Their test is similar to the above test for the last example. The last two papers report simulation results that support the superiority of these methods over a selected complete case method. However, one can verify that the first test statistic in Sun, Wang and Dai (2009) is asymptotically equivalent to a complete case statistic in their case 3, and this complete case statistic should thus result in an equivalent test. Finally, Li (2012) uses imputation together with the minimum distance methodology of Koul and Ni (2004) to propose tests for fitting a class of parametric models to the regression function that includes polynomials.

This article is organized as follows. Section 2 provides the theory for the transfer principle. The key is Lemma 2.1, which calculates the explicit form of the distribution of a complete case statistic. In Section 3 we show that the influence function of the complete case version of an efficient estimator of ϑ in the full data partially linear model is the efficient influence function for estimating ϑ in the MAR model. Similar results are sketched for a partially linear additive model (see Remark 3.1) and a single index model (see Remark 3.3). Section 4 discusses the test for normality of the errors for the MAR model, and derives expansions for the complete case residual-based empirical distribution function. In Section 5 we provide details for the complete case version of the second test about the nonparametric part ρ in (1.1) being constant.

2. Distribution theory for general complete case statistics. In this section we derive the exact distribution of a complete case statistic in a general setting. Let $(\mathcal{X}, \mathscr{A})$ be a measurable space, and, for each integer k, let t_k be a measurable function from $\mathcal{X}^k$ into $\mathbb{R}^m$. Let $(\delta_1, \xi_1), (\delta_2, \xi_2), \dots$ be independent copies of (δ, ξ), where δ is Bernoulli with parameter $p > 0$ and ξ is a $\mathcal{X}$-valued random variable. We denote the conditional distribution of ξ, given $\delta = 1$ by $\tilde{Q}$. Let $\tilde{\xi}_1, \tilde{\xi}_2, \dots$ be independent $\mathcal{X}$-valued random variables with common distribution $\tilde{Q}$. Denote the distribution of $t_n(\tilde{\xi}_1, \dots, \tilde{\xi}_n)$ by R_n. Then, for any Borel set B,

$$\begin{aligned} R_n(B) = \tilde{Q}^n(t_n \in B) &= P\big(t_n(\tilde{\xi}_1, \dots, \tilde{\xi}_n) \in B\big) \\ &= P\big(t_n(\xi_1, \dots, \xi_n) \in B | \delta_1 = 1, \dots, \delta_n = 1\big). \end{aligned}$$

By a *complete case statistic associated with the sequence* (t_n) we mean a statistic $T_{c,n}$ of the form

$$T_{c,n} = \sum_{A \subset \{1,\dots,n\}} \Big\{\prod_{i \in A} \delta_i\Big\}\Big\{\prod_{i \notin A} (1 - \delta_i)\Big\} t_{|A|}(\xi^A),$$

where $t_0(\xi^{\varnothing})$ is a constant, $|A|$ denotes the cardinality of A and ξ^A is the vector $(\xi_{i_1}, \dots, \xi_{i_k})$ with $i_1 < \dots < i_k$ the elements of the non-empty subset $A \subset \{1, \dots, n\}$. Note that the product $[\prod_{i \in A} \delta_i][\prod_{i \notin A}(1 - \delta_i)]$ is the indicator function of the event $\{\delta_i = 1, i \in A\} \cap \{\delta_i = 0, i \notin A\}$. Thus, $T_{c,n}$ equals $t_{|A|}(\xi^A)$ on this event. It is now clear that $T_{c,n}$ depends on the indicators $\delta_1, \dots, \delta_n$ and only those observations ξ_i for which $\delta_i = 1$.

REMARK 2.1. For a measurable function ψ defined on $\mathcal{X}$, we define the sequence $(\bar{\psi}_n)$ by $\bar{\psi}_n(x_1, \dots, x_n) = (\psi(x_1) + \dots + \psi(x_n))/n$. The complete case statistic associated with $(\bar{\psi}_n)$ is $\sum_{j=1}^n \delta_j \psi(\xi_j) / \sum_{j=1}^n \delta_j$.

REMARK 2.2. If $T_{c,n}$ is a complete case statistic associated with (t_n) and α is a real number, then $(\sum_{j=1}^n \delta_j)^\alpha T_{c,n}$ is a complete case statistic associated with the sequence $(n^\alpha t_n)$.

For the remainder of this section $T_{c,n}$ denotes a complete case statistic associated with (t_n) and H_n its distribution. The next lemma calculates H_n explicitly.

LEMMA 2.1. *For every Borel subset B of $\mathbb{R}^m$, we have*

$$H_n(B) = P(T_{c,n} \in B) = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} R_k(B),$$

with $R_0(B) = 1[t_0(\xi^{\varnothing}) \in B]$.

PROOF. Conditioning on $\delta_1, \dots, \delta_n$ yields the identity

$$P(T_{c,n} \in B) = E\big[P(T_{c,n} \in B|\delta_1, \dots, \delta_n)\big]$$

and, thus,

$$H_n(B) = \sum_{A \subset \{1,\dots,n\}} p^{|A|}(1-p)^{n-|A|} H(A, B),$$

where

$$\begin{aligned} H(A, B) &= P(T_{c,n} \in B|\delta_i = 1, i \in A, \delta_j = 0, j \notin A) \\ &= P\big(t_{|A|}(\xi^A) \in B|\delta_i = 1, i \in A, \delta_j = 0, j \notin A\big) \\ &= \tilde{Q}^{|A|}(t_{|A|} \in B) = R_{|A|}(B) \end{aligned}$$

for non-empty A, while $H(\varnothing, B) = R_0(B)$. The desired result is now immediate. □

REMARK 2.3. Lemma 2.1 has the following interpretation. The statistic $T_{c,n}$ has the same distribution as $t_K(\tilde{\xi}_1, \dots, \tilde{\xi}_K)$, where K is a binomial random variable with parameters n and p, independent of $\tilde{\xi}_1, \tilde{\xi}_2, \dots$.

From the lemma we immediately obtain the following results.

COROLLARY 2.1. *The following statements hold*:

(a) *If the sequence* (R_n) *is tight*, *so is the sequence* (H_n).
(b) *If* R_n *converges weakly to some limit* L, *then* H_n *converges weakly to the same limit* L.
(c) *If* R_n *converges weakly to point mass at* 0, *then* $T_{c,n}$ *converges in probability to zero*.

REMARK 2.4. Recall that R_n is the distribution of $t_n(\tilde{\xi}_1, \dots, \tilde{\xi}_n)$. Thus, by (b), the limiting distribution of $T_{c,n}$ equals the limiting distribution of $t_n(\tilde{\xi}_1, \dots, \tilde{\xi}_n)$. This provides the basis for the transfer principle.

REMARK 2.5. Let ψ and $\bar{\psi}_n$ be as in Remark 2.1. Set $N = \sum_{j=1}^n \delta_j$. Then

$$S_{c,n} = \sqrt{N}\left(T_{c,n} - \frac{1}{N}\sum_{j=1}^n \delta_j \psi(\xi_j)\right)$$

is a complete case statistic associated with $s_n = (n^{1/2}(t_n - \bar{\psi}_n))$. Suppose that

$$s_n(\tilde{\xi}_1, \dots, \tilde{\xi}_n) = n^{1/2}\left(t_n(\tilde{\xi}_1, \dots, \tilde{\xi}_n) - \frac{1}{n}\sum_{j=1}^n \psi(\tilde{\xi}_j)\right) = o_P(1).$$

Then, by (c), we have

$$S_{c,n} = \sqrt{N}\left(T_{c,n} - \frac{1}{N}\sum_{j=1}^{n} \delta_j \psi(\xi_j)\right) = o_P(1)$$

and, consequently,

$$T_{c,n} = \frac{1}{N}\sum_{j=1}^{n} \delta_j \psi(\xi_j) + o_P\big(n^{-1/2}\big).$$

This is the basis for the transfer principle for asymptotically linear statistics.

3. Efficiency considerations for the partially linear MAR model. Here we shall show that the expansion (1.2) characterizes efficient estimators in the partially linear MAR model. For this we only need to show that the influence function appearing in (1.2) is the efficient influence function for estimating ϑ in this model. We formulate this as the main result of this section; see Lemma 3.1. By the discussion in the Introduction, we must require that the conditional distribution $\tilde{G}$ of (U, V), given $\delta = 1$, satisfies the assumptions (G1) and (G2). This is crucial for the transfer principle to apply, and holds if the function π is bounded away from zero, as we shall show first.

REMARK 3.1. Consider the conditional distribution $\tilde{G}$ of (U, V) given $\delta = 1$. Then $\tilde{G}$ satisfies the properties (G1) and (G2) if π is bounded away from zero: it is easy to check that $\tilde{G}$ has density $\tilde{\pi}$ with respect to G, where $\tilde{\pi}(U, V) = \pi(U, V)/E[\delta]$. If $\tilde{\pi} \geq \eta$ for some positive constant η, then

$$\eta \int |h|\, dG \leq \int |h|\, d\tilde{G} \leq \int |h|\, dG/E[\delta]$$

for all $h \in L_1(G)$ and, therefore,

$$\begin{aligned} a^\top W_{\tilde{G}} a &= \int \big|a^\top\big(u - \mu_{\tilde{G}}(v)\big)\big|^2\, d\tilde{G}(u, v) \geq \eta \int \big|a^\top\big(u - \mu_{\tilde{G}}(v)\big)\big|^2\, dG(u, v) \\ &\geq \eta \int \big|a^\top\big(u - \mu_G(v)\big)\big|^2\, dG(u, v) = \eta a^\top W_G a \qquad \text{for all } a \in \mathbb{R}^m. \end{aligned}$$

From these inequalities we conclude that $\tilde{G}$ inherits the properties (G1) and (G2) from G if π is bounded away from zero.

LEMMA 3.1. *Suppose the model* (1.1) *holds with ρ being twice continuously differentiable and error density having finite Fisher information for location. Also assume π is bounded away from zero. Then an efficient estimator of the parameter ϑ in the MAR model is characterized by* (1.2). *As a consequence, the complete case version of an efficient estimator of the parameter ϑ in the full model is efficient for the MAR model.*

PROOF. We rely heavily on the calculations in Müller, Schick and Wefelmeyer (2006). The authors considered the general missing data problem with base observation $(X, \delta Y, \delta)$ where X and Y do not have to follow a regression model. They expressed the joint distribution P of $(X, \delta Y, \delta)$ via

$$P(dx, dy, dz) = G(dx) B_{\pi(x)}(dz)\big(z Q(x, dy) + (1-z)\Delta_0(dy)\big)$$

in terms of the distribution G of X, the conditional probability $\pi(x)$ of $\delta = 1$ given $X = x$, and the conditional distribution $Q(x, dy)$ of Y given $X = x$. Here B_p denotes the Bernoulli distribution with parameter p and Δ_t the Dirac measure at t. They showed that the tangent space is the sum of the orthogonal spaces

$$T_1 = \{u(X) : u \in \mathscr{U}\}, \qquad T_2 = \{\delta v(X, Y) : v \in \mathscr{V}\},$$
$$T_3 = \{(\delta - \pi(X)) w(X) : w \in \mathscr{W}\}.$$

Here, the set $\mathscr{U}$ consists of all real-valued functions u satisfying $\int u\, dG = 0$, $\int u^2\, dG < \infty$ and for which there is a sequence G_{nu} of distributions fulfilling the model assumptions on G and

$$\int \left(n^{1/2}\big(dG_{nu}^{1/2} - dG^{1/2}\big) - \frac{1}{2} u\, dG^{1/2} \right)^2 \to 0.$$

The set $\mathscr{W}$ consists of real-valued functions w with the property $\int w^2 \pi(1 - \pi)\, dG < \infty$ for which there is a sequence π_{nw} satisfying the model assumptions on π such that

$$\int \left(n^{1/2}\big(dB_{\pi_{nw}(x)}^{1/2} - dB_{\pi(x)}^{1/2}\big) - \frac{1}{2}\big(\cdot - \pi(x)\big)\, dB_{\pi(x)}^{1/2} \right)^2 G(dx) \to 0.$$

Finally, the set $\mathscr{V}$ consists of functions v with the properties $\int v(x, y) Q(x, dy) = 0$ for all x and $\int v^2(x, y) G(dx) Q(x, dy) < \infty$, and for which there is a sequence Q_{nv} satisfying the model assumptions on Q and

$$\iint \left(n^{1/2}\big(dQ_{nv}^{1/2}(x, \cdot) - dQ^{1/2}(x, \cdot)\big) - \frac{1}{2} v(x, \cdot)\, dQ^{1/2}(x, \cdot) \right)^2 G(dx) \to 0.$$

In the partially linear regression model (1.1) we have $X = (U, V)$ and

$$Q(x, dy) = Q_{\vartheta, \rho, f}(u, v, dy) = f\big(y - \vartheta^\top u - \rho(v)\big)\, dy,$$

where the density f has finite Fisher information for location, ϑ belongs to $\mathbb{R}^m$ and ρ is a smooth function. For this model $\mathscr{V}$ consists of the functions

$$a^\top U \ell_f(\varepsilon) + b(V)\ell_f(\varepsilon) + c(\varepsilon)$$

with $a \in \mathbb{R}^m$, $E[b^2(V)] < \infty$ and $c \in L_2(F)$ with $\int c(y)\, dF(y) = 0$ and $\int c(y) y\, dF(y) = 0$. Since we are interested in estimating the finite-dimensional parameter ϑ, we introduce the functional

$$\kappa(G, Q_{\vartheta, \rho, f}, \pi) = \vartheta.$$

Now consider

$$g(X, \delta Y, \delta) = \delta\big(U - \mu_1(V)\big)\ell_f(\varepsilon)$$

with $\mu_1(V) = E(U|V, \delta = 1)$. Then the coordinates of $g(X, \delta Y, \delta)$ belong to $\mathscr{V}$. Thus, we have $E[g(X, \delta Y, \delta)u(X)] = 0$ and $E[g(X, \delta Y, \delta)(\delta - \pi(X))w(X)] = 0$. Note that ε and (δ, X) are independent, and that we have $E[\ell_f(\varepsilon)] = 0$ and $E[\ell_f^2(\varepsilon)] = J_f$. Using this and the definition of μ_1, we calculate

$$\begin{aligned}
&E\big[g(X, \delta Y, \delta)\delta\big(a^\top U\ell_f(\varepsilon) + b(V)\ell_f(\varepsilon) + c(\varepsilon)\big)\big] \\
&\quad = E\big[\delta\big(U - \mu_1(V)\big)\big(U^\top a + b(V)\big)\big]J_f + E\big[\delta\big(U - \mu_1(V)\big)\big]E\big[\ell_f(\varepsilon)c(\varepsilon)\big] \\
&\quad = E\big[\delta\big(U - \mu_1(V)\big)\big(U - \mu_1(V)\big)^\top\big]aJ_f.
\end{aligned}$$

From this we can conclude that the functional κ is differentiable with canonical gradient $g_*(X, \delta Y, \delta)$ of the form

$$\begin{aligned}
&\delta\big(J_f E\big[\delta\big(U - \mu_1(V)\big)\big(U - \mu_1(V)\big)^\top\big]\big)^{-1}\big(U - \mu_1(V)\big)\ell_f(\varepsilon) \\
&\quad = \frac{\delta}{E[\delta]}\big(J_f E\big[\big(U - \mu_1(V)\big)\big(U - \mu_1(V)\big)^\top|\delta = 1\big]\big)^{-1}\big(U - \mu_1(V)\big)\ell_f(\varepsilon).
\end{aligned}$$

This canonical gradient is the influence function of an efficient estimator of ϑ. Now use the fact that $\mu_1(V)$ equals $\mu_{\tilde{G}}(V)$ and $E[(U - \mu_1(V))(U - \mu_1(V))^\top|\delta = 1]$ equals $W_{\tilde{G}}$ to see that this is indeed the characterization (1.2). $\square$

REMARK 3.2. The above efficiency result extends in a straightforward manner to the case when V is higher dimensional. It also extends to the partially linear additive model

$$Y = \vartheta^\top U + \rho_1(V_1) + \rho_2(V_2) + \varepsilon,$$

where (V_1, V_2) takes values in the unit square $[0, 1]^2$ and has a density that is bounded and bounded away from zero on the unit square. Let G now denote the joint distribution of (U, V_1, V_2). Assume that the matrix $E[(U - \mu_G(V_1, V_2))(U - \mu_G(V_1, V_2))]^\top$, with $\mu_G(V_1, V_2) = E(U|(V_1, V_2))$, is positive definite, and that π is bounded away from zero. In the present model the space $\mathscr{V}_1$ consists of functions of the form

$$a^\top U\ell_f(\varepsilon) + \big(b_1(V_1) + b_2(V_2)\big)\ell_f(\varepsilon) + c(\varepsilon),$$

where $E[b_1^2(V_1) + b_2^2(V_2)]$ is finite. The role of g is now played by

$$g(X, \delta Y, \delta) = \delta\big(U - \tilde{\nu}_1(V_1) - \tilde{\nu}_2(V_2)\big)\ell_f(\varepsilon),$$

where $\tilde{\nu}_1(V_1) + \tilde{\nu}_2(V_2)$ minimizes $E[\|U - B_1(V_1) - B_2(V_2)\|^2|\delta = 1]$ with respect to functions B_1 and B_2 from $[0, 1]$ into $\mathbb{R}^m$ such that $E[\|B_1(V_1)\|^2]$ and

$E[\|B_2(V_2)\|^2]$ are finite. The efficient influence function is

$$\frac{\delta}{E[\delta]}\big(J_f E\big[\big(U-\tilde{\nu}_1(V_1)-\tilde{\nu}_2(V_2)\big)\big(U-\tilde{\nu}_1(V_1)-\tilde{\nu}_2(V_2)\big)^\top|\delta=1\big]\big)^{-1} \times \big(U-\tilde{\nu}_1(V_1)-\tilde{\nu}_2(V_2)\big)\ell_f(\varepsilon).$$

By the transfer principle, this is the influence function of a complete case version of an estimator with influence function

$$\big(J_f E\big[\big(U-\nu_1(V_1)-\nu_2(V_2)\big)\big(U-\nu_1(V_1)-\nu_2(V_2)\big)^\top\big]\big)^{-1} \times \big(U-\nu_1(V_1)-\nu_2(V_2)\big)\ell_f(\varepsilon)$$

in the full model, where $\nu_1(V_1)+\nu_2(V_2)$ minimizes $E[\|U-B_1(V_1)-B_2(V_2)\|^2]$ over functions B_1 and B_2 as above. Schick (1996) constructed estimators in the full model that have the latter influence function. In particular, he established their efficiency by showing that the above influence function is indeed the efficient influence function in the full model.

REMARK 3.3. In the above we have shown that for the partially linear MAR model (with a possibly additive smooth function) an efficient estimator of the parameter ϑ can be obtained as the complete case version of an efficient estimator in the full model. This is typically also true for other more general semiparametric regression models and can be verified along the above lines. We sketch this for the following single index model.

In this model $Y=\rho(V+\vartheta^\top U)+\varepsilon$ with one-dimensional V, m-dimensional U, $\vartheta\in\mathbb{R}^m$ and twice continuously differentiable ρ. Assume again that π is bounded away from zero. The space $\mathscr{V}$ for this model consists of functions

$$a^\top U\rho'\big(V+\vartheta^\top U\big)\ell_f(\varepsilon)+b\big(V+\vartheta^\top U\big)\ell_f(\varepsilon)+c(\varepsilon)$$

with $a\in\mathbb{R}^m$, $E[b^2(V+\vartheta^\top U)]<\infty$ and c as before. For this we must require that $E[\|U\|^2(\rho'(V+\vartheta^\top U))^2]$ is finite. Now one works with $g(X,\delta Y,\delta)=\delta(U-\nu_1(V+\vartheta^\top U))\rho'(V+\vartheta^\top U)\ell_f(\varepsilon)$ and $\nu_1(V+\vartheta^\top U)=E(U|V+\vartheta^\top U,\delta=1)$, and obtains the canonical gradient

$$g_*(X,\delta Y,\delta)=\frac{\delta}{E[\delta]}(J_f W_1)^{-1}\big(U-\nu_1\big(V+\vartheta^\top U\big)\big)\rho'\big(V+\vartheta^\top U\big)\ell_f(\varepsilon)$$

if $W_1=E[(U-\nu_1(V+\vartheta^\top U))(U-\nu_1(V+\vartheta^\top U))^\top(\rho'(V+\vartheta^\top U))^2|\delta=1]$ is invertible. By the transfer principle, this is the influence function of a complete case version of an estimator with influence function $(J_f W)^{-1}(U-\nu(V+\vartheta^\top U))\rho'(V+\vartheta^\top U)\ell_f(\varepsilon)$, where $\nu(V+\vartheta^\top U)=E[U|V+\vartheta^\top V]$ and $W=E[(U-\nu(V+\vartheta^\top U))(U-\nu(V+\vartheta^\top U))^\top](\rho'(V+\vartheta^\top U))^2]$. The latter influence function is the efficient gradient for the full model. Indeed, it is the canonical gradient for the case when $\delta=1$ almost surely.

4. Testing for normal errors. In this section we shall introduce a test for normal errors which uses the Khmaladze transform of the empirical distribution function $\hat{\mathbb{F}}(t) = n^{-1}\sum_{j=1}^n \mathbf{1}[\hat\varepsilon_j \le t]$, $t \in \mathbb{R}$, based on residuals $\hat\varepsilon_j$. Goodness-of-fit tests for the full model based on that transform were discussed in Khmaladze and Koul (2004, 2009) for parametric and nonparametric regression, and by MSW for the partially linear regression model considered here. Due to the transfer principle, it is now straightforward to adapt the approach by MSW to the MAR model, which is what we will do here for a simple illustration of the method. Note that MSW consider the more complex case where V is a covariate *vector*.

First, we briefly sketch the approach for the full model. To avoid additional assumptions, we estimate ϑ and ρ using a least squares approach with the trigonometric basis. This is discussed in MSW, Section 4, for an additive regression function, that is, with $\rho(x_1, \dots, x_q) = \rho_1(x_1) + \cdots + \rho_q(x_q)$. (Here we have $q = 1$.) For $k = 1, 2, \dots,$ we set

$$\phi_k(x) = \cos(\pi k x), \qquad 0 \le x \le 1.$$

Our estimator of the regression function $r(u, v) = \vartheta^\top u + \rho(v)$ is then

$$\hat r(u, v) = \hat\vartheta^\top u + \sum_{k=0}^K \hat\beta_k \phi_k(v),$$

where $\phi_0(x) = 1$ and $(\hat\vartheta^\top, \hat\beta_0, \dots, \hat\beta_K)$ minimizes

$$\sum_{j=1}^n \left(Y_j - a^\top U_j - \sum_{k=0}^K b_k \phi_k(V_j) \right)^2$$

with respect to $a, b_0, \dots, b_K$. For $j = 1, \dots, n$ the error ε_j is estimated by the residual

$$\hat\varepsilon_j = Y_j - \hat\vartheta^\top U_j - \sum_{k=0}^K \hat\beta_k \phi_k(V_j).$$

We also need the normalized residuals $\hat Z_j = \hat\varepsilon_j / \hat\sigma$, where $\hat\sigma$ is the square root of $(1/n)\sum_{j=1}^n \hat\varepsilon_j^2$.

Assume for the remainder of this section that f has finite Fisher information for location and finite fourth moment. This assumption is met by the normal density. It then follows from MSW, Theorem 4.1 and Remark 4.2, that with $K = K_n \sim n^{-1/4}$ we have the uniform stochastic expansions

$$\sup_{t\in\mathbb{R}} \left| \frac{1}{\sqrt n} \sum_{j=1}^n \big(\mathbf{1}[\hat\varepsilon_j \le t] - \mathbf{1}[\varepsilon_j \le t] - f(t)\varepsilon_j \big) \right| = o_P(1) \tag{4.1}$$

and

$$\sup_{t\in\mathbb{R}} \left| \frac{1}{\sqrt n} \sum_{j=1}^n \left(\mathbf{1}[\hat Z_j \le t] - \mathbf{1}[Z_j \le t] - f_*(t)\left(Z_j + t\frac{Z_j^2 - 1}{2} \right) \right) \right| = o_P(1), \tag{4.2}$$

where f_* denotes the density of the normalized errors $Z_j = \varepsilon_j/\sigma$.

Write ϕ for the standard normal density. In terms of the density f_*, the null hypothesis is

$$H_0 : f_* = \phi.$$

MSW proposed the test statistic

$$T_n = \sup_{t\in\mathbb{R}} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \big(\mathbf{1}[\hat{Z}_j \le t] - H(t \wedge \hat{Z}_j) h(\hat{Z}_j)\big) \right|,$$

with

$$\text{(4.3)} \qquad \begin{aligned} h(x) &= \big(1, x, x^2 - 1\big)^\top, \qquad \Gamma(x) = \int_x^\infty h(z)h^\top(z)\phi(z)\,dz, \\ H(t) &= \int_{-\infty}^{t} h^\top(x)\Gamma^{-1}(x)\phi(x)\,dx. \end{aligned}$$

This is a version of the martingale transform test of Khmaladze and Koul (2009) for fitting an error distribution in nonparametric regression. MSW showed that under the null hypothesis the test statistic T_n converges in distribution to ζ, which is the supremum of a standard Brownian motion given in (1.3). This holds for every distribution function G satisfying (G1) and (G2). Since ε and (δ, U, V) are independent under the MAR assumption, the conditional distribution of (ε, U, V), given $\delta = 1$, is given by $F \times \tilde{G}$, where $\tilde{G}$ is the conditional distribution of (U, V), given $\delta = 1$. Thus, if $\tilde{G}$ satisfies (G1) and (G2), then the transfer principle applies and yields the same limiting distribution for the complete case version T_c of T_n, where

$$\text{(4.4)} \qquad T_c = \sup_{t\in\mathbb{R}} \left| \frac{1}{\sqrt{N}} \sum_{j=1}^{n} \delta_j \big(\mathbf{1}[\hat{Z}_{jc} \le t] - H(t \wedge \hat{Z}_{jc}) h(\hat{Z}_{jc})\big) \right|.$$

Here $\hat{Z}_{jc}$ are the complete case versions of the normalized residuals and are defined by $\hat{\varepsilon}_{jc}/\hat{\sigma}_c$ with $\hat{\varepsilon}_{jc} = Y_j - \hat{\vartheta}_c^\top U_j - \sum_{k=0}^{K_N} \hat{\beta}_k \psi_k(V_j)$ and $\hat{\sigma}_c$ the square root of $N^{-1}\sum_{j=1}^{n} \delta_j \hat{\varepsilon}_{jc}^2$, while $(\hat{\vartheta}_c^\top, \hat{\beta}_0, \dots, \hat{\beta}_{K_N})$ are the least squares estimators minimizing

$$\sum_{j=1}^{n} \delta_j \left(Y_j - a^\top U_j - \sum_{k=0}^{K_N} b_k \phi_k(V_j) \right)^2.$$

The transfer principle for asymptotically linear statistics also provides complete case versions of the expansions (4.1) and (4.2) from above. The first expansion becomes

$$\sup_{t\in\mathbb{R}} \left| \frac{1}{\sqrt{N}} \sum_{j=1}^{n} \delta_j \big(\mathbf{1}[\hat{\varepsilon}_{jc} \le t] - \mathbf{1}[\varepsilon_j \le t] - f(t)\varepsilon_j\big) \right| = o_P(1),$$

and the second expansion becomes

$$\sup_{t\in\mathbb{R}}\left|\frac{1}{\sqrt{N}}\sum_{j=1}^{n}\delta_j\left(\mathbf{1}[\hat{Z}_{jc}\le t]-\mathbf{1}[Z_j\le t]-f_*(t)\left(Z_j+t\frac{Z_j^2-1}{2}\right)\right)\right|=o_P(1).$$

5. Testing for linearity. In this section we address testing whether the function ρ in the partially linear MAR model is constant. In the previous section we demonstrated how the transfer principle can be used to adapt a known test for the full model to the MAR model. We now show how to develop a test procedure for the MAR model when no counterpart to the full model exists. Our approach is to first develop a procedure for the full model, and then to apply the transfer principle. Our test statistic is inspired by that in Stute, Xu and Zhu (2008).

Under the null hypothesis the partially linear model reduces to the linear regression model $Y=\alpha+\vartheta^\top U+\varepsilon$, where α is an unknown constant, that is, we have

$$H_0:\rho(v)=\alpha \qquad \text{for all } v\in\mathbb{R} \text{ and some } \alpha\in\mathbb{R}.$$

To simplify notation, we introduce $\beta=(\alpha,\vartheta^\top)^\top$ and $Z=(1,U^\top)^\top$. Then we can write the model under the null hypothesis as $Y=\beta^\top Z+\varepsilon$.

It follows from (G2) that the dispersion matrix Λ_G of U is positive definite. From this we immediately see that the matrix

$$M_G=E[ZZ^\top]=\begin{bmatrix}1 & E[U^\top]\\ E[U] & E[UU^\top]\end{bmatrix}$$

is also positive definite. Thus, the least squares estimator $\hat{\beta}$ of $\beta=(\alpha,\vartheta^\top)^\top$ is root-n consistent under the null hypothesis, as it satisfies

$$\hat{\beta}=\beta+M_G^{-1}\frac{1}{n}\sum_{j=1}^{n}Z_j\varepsilon_j+o_P(n^{-1/2}).$$

Now let χ denote a continuous non-constant function on [0, 1]. Introduce the least squares estimator $\hat{\gamma}$ for regressing the responses $\chi(V_j)$ on the design vectors Z_j, so that $\hat{\gamma}$ minimizes

$$\frac{1}{n}\sum_{j=1}^{n}(\chi(V_j)-\gamma^\top Z_j)^2.$$

Set $R_j=\chi(V_j)-\hat{\gamma}^\top Z_j$, $W_j=R_j/(n^{-1}\sum_{j=1}^{n}R_j^2)^{1/2}$, and $\hat{\varepsilon}_{j0}=Y_j-\hat{\beta}^\top Z_j$, $j=1,\dots,n$. Our test statistic in the full model is

$$T_n=\sup_{t\in\mathbb{R}}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}W_j\mathbf{1}[\hat{\varepsilon}_{j0}\le t]\right|.$$

As in Stute, Xu and Zhu (2008), we have the following result.

LEMMA 5.1. *Suppose the null hypothesis holds and f is uniformly continuous. Then T_n converges in distribution to $\zeta_0=\sup_{0\le t\le 1}|B_0(t)|$, where B_0 denotes a standard Brownian bridge.*

PROOF. Set

$$\chi_G(X) = \chi(V) - \gamma_G^\top Z,$$

where γ_G minimizes $E[(\chi(V) - \gamma^\top Z)^2]$. Let $\rho_G = \Lambda_G^{-1} E[\chi(V)(U - E[U])]$. Then it is easy to check that

$$\begin{aligned}\chi_G(X) &= \chi(V) - E[\chi(V)] - \rho_G^\top(U - E[U]) \\ &= \chi(V) - E[\chi(V)] - \rho_G^\top(\mu_G(V) - E[U]) - \rho_G^\top(U - \mu_G(V)).\end{aligned}$$

Note also that χ being non-constant on $[0, 1]$ and V having a positive density on $[0, 1]$ implies $\chi(V)$ has a positive variance. These facts together with W_G being positive definite guarantee that $E[\chi_G^2(X)] = \mathrm{Var}(\chi(V) - \rho_G^\top \mu_G(V)) + \rho_G^\top W_G \rho_G$ is positive.

Next, let g be a measurable function such that $E[g^2(X)]$ is finite and assume f is uniformly continuous. Then Theorem 2.2.4 of Koul (2002) yields

$$\sup_{t\in\mathbb{R}}\left|\frac{1}{n}\sum_{j=1}^n g(X_j)(\mathbf{1}[\hat\varepsilon_{j0} \le t] - \mathbf{1}[\varepsilon_j \le t]) - f(t)E[g(X)Z^\top](\hat\beta - \beta)\right| = o_P(n^{-1/2}).$$

From this fact we obtain

$$\sup_{t\in\mathbb{R}}\left|\frac{1}{n}\sum_{j=1}^n R_j(\mathbf{1}[\hat\varepsilon_{j0} \le t] - \mathbf{1}[\varepsilon_j \le t]) - f(t)\hat D(\hat\beta - \beta)\right| = o_P(n^{-1/2}),$$

where

$$\hat D = E[\chi(V)Z^\top] - \hat\gamma^\top E[ZZ^\top] = E[\chi_G(V)Z^\top] + o_P(1).$$

In view of the identities $E[\chi_G(V)Z^\top] = 0$ and $\sum_{j=1}^n R_j = 0$, we can conclude

$$\sup_{t\in\mathbb{R}}\left|\frac{1}{n}\sum_{j=1}^n R_j\mathbf{1}[\hat\varepsilon_{j0} \le t] - \frac{1}{n}\sum_{j=1}^n R_j(\mathbf{1}[\varepsilon_j \le t] - F(t))\right| = o_P(n^{-1/2}).$$

Writing $R_j - \chi_G(V_j) = -(\hat\gamma - \gamma_G)^\top Z_j$, we derive the expansions

$$\sup_{t\in\mathbb{R}}\left|\frac{1}{n}\sum_{j=1}^n (R_j - \chi_G(V_j))(\mathbf{1}[\varepsilon_j \le t] - F(t))\right| = o_P(n^{-1/2}),$$

$$\frac{1}{n}\sum_{j=1}^n (R_j - \chi_g(V_j))^2 \le \|\hat\gamma - \gamma_G\|^2 \frac{1}{n}\sum_{j=1}^n \|Z_j\|^2 = o_P(1),$$

and therefore obtain $n^{-1}\sum_{j=1}^n R_j^2 = E[\chi_G^2(V)] + o_P(1)$. The above derivations in turn yield

$$\sup_{t\in\mathbb{R}}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^n W_j\mathbf{1}[\hat\varepsilon_{j0} \le t] - \frac{1}{\sqrt{n}}\sum_{j=1}^n \chi_G^*(V_j)(\mathbf{1}[\varepsilon_j \le t] - F(t))\right| = o_P(1),$$

with $\chi_G^* = \chi_G / E[\chi_G^2(V)]^{1/2}$. Since, again by Theorem 2.2.4 of Koul (2002), the process

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \chi_G^*(V_j)\big(\mathbf{1}[\varepsilon_j \le t] - F(t)\big), \qquad -\infty \le t \le \infty,$$

converges in $D([-\infty, \infty])$ to a time-changed Brownian bridge $B_0(F)$, we conclude that T_n has the desired limiting distribution. □

The complete case version of T_n is given by

$$T_c = \sup_{t\in\mathbb{R}} \left| \frac{1}{\sqrt{N}} \sum_{j=1}^{n} \delta_j R_{jc} \mathbf{1}[\hat{\varepsilon}_{jc} \le t] \right| \Big/ \left(\frac{1}{N} \sum_{j=1}^{n} \delta_j R_{jc}^2 \right)^{1/2}, \tag{5.1}$$

with $\hat{\varepsilon}_{jc} = Y_j - \hat{\beta}_c^\top Z_j$, $R_{jc} = \chi(V_j) - \hat{\gamma}_c^\top Z_j$,

$$\hat{\beta}_c = \arg\min_b \sum_{j=1}^{n} \delta_j \big(Y_j - b^\top Z_j\big)^2 \quad \text{and}$$

$$\hat{\gamma}_c = \arg\min_\gamma \sum_{j=1}^{n} \delta_j \big(\chi(V_j) - \gamma^\top Z_j\big)^2.$$

By the transfer principle, the limiting distribution of T_c under the null hypothesis will be that of ζ_0 from the above lemma, as long as f is uniformly continuous and $\tilde{G}$ satisfies (G1) and (G2).

REMARK 5.1. The above is easily extended to cover testing for other parametric forms for ρ. For example, we can test whether ρ is linear, $\rho(v) = a + bv$. In this case we proceed as above, but with the role of Z now played by the vector $(1, U^\top, V)$ and with χ chosen to be nonlinear.

Acknowledgments. The authors would like to thank the two reviewers for their constructive comments, which have helped to improve the paper.

REFERENCES

BHATTACHARYA, P. K. and ZHAO, P.-L. (1997). Semiparametric inference in a partial linear model. *Ann. Statist.* **25** 244–262. MR1429924

CUZICK, J. (1992). Efficient estimates in semiparametric additive regression models with unknown error distribution. *Ann. Statist.* **20** 1129–1136. MR1165611

EFROMOVICH, S. (2011). Nonparametric regression with responses missing at random. *J. Statist. Plann. Inference* **141** 3744–3752. MR2823645

FORRESTER, J., HOOPER, W., PENG, H. and SCHICK, A. (2003). On the construction of efficient estimators in semiparametric models. *Statist. Decisions* **21** 109–137. MR2000666

GONZÁLEZ-MANTEIGA, W. and PÉREZ-GONZÁLEZ, A. (2006). Goodness-of-fit tests for linear regression models with missing response data. *Canad. J. Statist.* **34** 149–170. MR2280923

HART, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer, New York. MR1461272

KHMALADZE, E. V. and KOUL, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.* **32** 995–1034. MR2065196

KHMALADZE, E. V. and KOUL, H. L. (2009). Goodness-of-fit problem for errors in nonparametric regression: Distribution free approach. *Ann. Statist.* **37** 3165–3185. MR2549556

KOUL, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*. *Lecture Notes in Statistics* **166**. Springer, New York. MR1911855

KOUL, H. L. (2006). Model diagnostics via martingale transforms: A brief review. In *Frontiers in Statistics* (J. Fan and H. L. Koul, eds.) 183–206. Imp. Coll. Press, London. MR2326002

KOUL, H. L. and NI, P. (2004). Minimum distance regression model checking. *J. Statist. Plann. Inference* **119** 109–141. MR2018453

LI, X. (2012). Lack-of-fit testing of a regression model with response missing at random. *J. Statist. Plann. Inference* **142** 155–170. MR2827138

LITTLE, R. J. A. and RUBIN, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. Wiley, Hoboken, NJ. MR1925014

MÜLLER, U. U. (2009). Estimating linear functionals in nonlinear regression with responses missing at random. *Ann. Statist.* **37** 2245–2277. MR2543691

MÜLLER, U. U., SCHICK, A. and WEFELMEYER, W. (2006). Imputing responses that are not missing. In *Probability, Statistics and Modelling in Public Health* (M. Nikulin, D. Commenges and C. Huber, eds.) 350–363. Springer, New York. MR2230741

MÜLLER, U. U., SCHICK, A. and WEFELMEYER, W. (2012). Estimating the error distribution function in semiparametric additive regression models. *J. Statist. Plann. Inference* **142** 552–566. MR2843057

SCHICK, A. (1993). On efficient estimation in regression models. *Ann. Statist.* **21** 1486–1521. MR1241276

SCHICK, A. (1996). Root-n-consistent and efficient estimation in semiparametric additive regression models. *Statist. Probab. Lett.* **30** 45–51. MR1411180

STUTE, W., XU, W. L. and ZHU, L. X. (2008). Model diagnosis for parametric regression in high-dimensional spaces. *Biometrika* **95** 451–467. MR2521592

SUN, Z. and WANG, Q. (2009). Checking the adequacy of a general linear model with responses missing at random. *J. Statist. Plann. Inference* **139** 3588–3604. MR2549107

SUN, Z., WANG, Q. and DAI, P. (2009). Model checking for partially linear models with missing responses at random. *J. Multivariate Anal.* **100** 636–651. MR2478187

H. L. KOUL
DEPARTMENT OF STATISTICS AND PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824-1027
USA
E-MAIL: koul@stt.msu.edu
URL: http://www.stt.msu.edu/~koul

U. U. MÜLLER
DEPARTMENT OF STATISTICS
TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843-3143
USA
E-MAIL: uschi@stat.tamu.edu
URL: http://www.stat.tamu.edu/~uschi

A. SCHICK
DEPARTMENT OF MATHEMATICAL SCIENCES
BINGHAMTON UNIVERSITY
BINGHAMTON, NEW YORK 13902-6000
USA
E-MAIL: anton@math.binghamton.edu
URL: http://www.math.binghamton.edu/anton