Vol. 00, No. 00, Month 200x, 1–13

RESEARCH ARTICLE

Efficient Inference in Semiparametric Generalized Linear Models

Hanxiang Peng*
Department of Mathematics, University of Mississippi
University, MS 38677, USA

(*Received 00 Month 200x; in final form 00 Month 200x*)

In this article, we calculate an explicit formula of the efficient influence function for a semiparametric generalized linear model using the method of orthogonality calculations. We construct an efficient estimate of the parameter of interest.

Keywords: Efficient influence function; Generalized linear model; Kernel estimate; Moment estimate; Semiparametric regression.

AMS Subject Classification: Primary 62G05, 62G20.

1. Introduction.

In a multivariate semiparametric generalized linear model (SGLM), the response $Y \in \ ^l$ and the covariate $(X, Z) \in [c, d]^{p\times m} \times [0, 1]$ satisfy the structural relation that the conditional expectation of Y given (X, Z) is related to a semiparametric systematic part $X^\top\vartheta + \varrho(Z)$ through a link h. Namely,

$$(Y|X, Z) = h(X^\top\vartheta + \varrho(Z)), \quad \vartheta \in \Theta, \quad \varrho \in \Upsilon, \tag{1}$$

where $\Theta \subset \ ^p$ is a nonempty open subset, Υ is a collection of unknown smooth functions from $[0, 1]$ into $\ ^m$, and h is a smooth function from $\ ^m$ to $\ ^l$. The conditional density f of Y of the conditional distribution F given (X, Z) w.r.t. a σ-finite measure λ_1 is assumed to be from an exponential family of the form

$$f(y|\varphi) = \exp(\varphi^\top y - b(\varphi)), \quad y \in \mathcal{Y}, \quad \varphi \in \Phi$$

for $\mathcal{Y}$ and Φ nonempty subsets of $\ ^l$. The covariate (X, Z) has an unknown joint distribution $G \in \mathcal{G}$, a nonempty collection of distributions.

Suppose from now on that the true but unknown parameters are (θ, ρ). We are interested in the *efficient* estimation of the regression parameter θ in the presence of the nuisance parameter $\gamma = (G, \rho)$ based on the independent and identically distributed observations $\xi_j = (X_j, Z_j, Y_j), j = 1, \cdots, n$ of $\xi = (X, Z, Y)$. The *efficiency* criterion used here is that of *least dispersed regular estimates* that is based on the convolution theorems. See, Begun, Hall and Wellner (1983), Pfanzagl and Wefelmeyer (1982), Bickel, Klaassen, Ritov and Wellner (1993) among others.

Relatively recently, a theory was developed on construction of efficient estimates based on the modern asymptotic theory of semiparametric models. Bickel (1982)

*Corresponding author. Email: mmpeng@olemiss.edu
ISSN: 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/1048525YYxxxxxxxx
http://www.informaworld.com

used the sample splitting technique to give a general procedure for constructing adaptive estimates in semiparametric models. The sample splitting was furthered in a series of articles by Schick (1986, 1987, 1993, 2001) and Forrester, *et al.* (2003) among others. The procedure nowadays utilizes all the data in both estimating the unknown influence function and evaluating the estimate functions, and thus has reached a stage of practical applications. In the spirit of this procedure, we construct an efficient estimate of the regression parameter based on all the sample in both estimation and evaluation. We avoid the sample splitting and discretization of the preliminary estimate. Discretization was introduced by Le Cam (1955) and is commonly used to circumvent technical difficulties in the construction of efficient estimates in modern semiparametric models. Discretized estimates can be used as if they were nonrandom thus simplify proofs.

The SGLMs lack the structural independence between the response and the unobservable error; which is usually assumed in various models such as the symmetric regression model, the partially linear model (PLM). The independence structure makes it possible for the conditions with the "leave-one-out" technique to be circumvented by conditioning argument. See Schick (1994, 2001) and Forrester, *et al.* (2003). However, no such assumption is allowed here and so we have to directly deal with this difficulty. We believe that our result is a supplement to the aforementioned work. It provides a theoretical justification for the commonly used *scoring method* about improving the efficiency of estimates of parameters in SGLMs. It sheds additional light on the existing theory of the construction of efficient estimates in semiparametric models.

The SGLM is a natural extension of the generalized linear models (GLM) studied by Nelder and Wedderburn (1972). When the nonparametric part ρ is left out, the SGLM recovers the GLM,

$$\mu = \quad (Y|X) = h(X^\top \theta).$$

When h is the identity map on , the SGLM reproduces the PLM, which is widely studied in literature, e.g. Schick (1993), Bhattacharya and Zhao (1997), and Forrester, *et al.* (2003) among others. Many authors have investigated *efficient* estimation of the parameters in the GLMs and SGLMs. Mammen and Geer (1997) studied quasilikelihood generalized linear models(QSLMs) using penalized quasi-likelihood method. They pointed out in Remark 4.2 (in the middle of page 1026) that their estimate is *efficient* if the conditional distribution of Y belongs to an exponential family with mean μ and variance $V(\mu)$. Chen (1995) considered the estimation of the SGLM when the link is canonical and proposed his estimate by utilizing regression splines in which the nonparametric part (ρ in our notation) was approximated by parametrized functions (linear combination of certain tensor-product polynomial splines). He indicated in Remark 1 (page 1109) that his estimate is *efficient* when the nonparametric ρ is approximated by parametrized functions of linear combination of certain tensor-product polynomial splines. Severini and Wong (1992) gave a quite general procedure, the *profile likelihood*, for efficient estimates for a class of semi-parametric models of which the generalized linear models follow as a special case. The profile likelihood estimates are only *efficient* provided that $\rho(z;\theta)$ is a "least favorable curve". Sevrini and Staniswalis (1994) investigated quasi-likelihood estimation in semiparametric models taking the above approach of "maximum likelihood estimation".

The formula of the *efficient influence function* for the SGLMs can be identified from different approaches. For example, those procedures given by Mammen and Geer (1997), Chen (1995), and Severini and Staniswalis (1992). Employing a standard technique in deriving efficient influence functions in semiparametric models, the method of *orthogonality calculations*, we calculate the explicit formula (cf., (5) and (6)). For further discussion and additional examples of the orthogonal calculations, see, the monograph by Bickel, *et al.* (1993). Our proposed efficient estimate is given in an explicit formula based on a preliminary estimate and is very convenient to manipulate, whereas estimates related to maximum likelihood such as those noted above usually involve in the solutions of highly nonlinear equations. With our method, we can readily improve an inefficient estimator to attain the efficiency.

In the sequel, we write θ, ρ for the true but unknown parameter values and ϑ, ϱ for the generic values in Θ, Υ respectively. Write ${}_{\vartheta,\gamma}$ the expectation calculated under the probability measure $P_{\vartheta,\gamma}$ and $P = P_{\theta,\gamma}$ and $\;\; = {}_{\theta,\gamma}$. For a probability measure Q, $L_{2,0}(Q) \equiv \{h \in L_2(P) : \int h\,dP = 0\}$. We call (and denote) $\{\theta_n : n = 1, 2, \cdots\} \subset \Theta$ a *local sequence* of θ if $\{\sqrt{n}(\theta_n - \theta)\}$ stays bounded uniformly in n. At the true parameters (θ, γ), we shall reserve $\eta = X^\top\theta + \rho(Z)$, $\phi = a(\eta)$; at the perturbed parameter value $\vartheta \in \Theta$, we reserve $\zeta = X^\top\vartheta + \rho(Z), \varphi = a(\zeta)$. We abuse φ to also denote an arbitrary element in Φ when there would be no ambiguity in the context. The rest of article is structured as follows. We calculate the information lower bound and characterize efficient estimates in Section 2. In Section 3, we propose an efficient estimate, followed by the main theorem. Section 4 deals with the convergence of the estimate of the nonparametric part and provides preliminary estimates. Section 5 collects technical details.

2. Efficiency Consideration

This section gives the efficient influence function via the method of orthogonality calculations and characterizes efficient regular estimates.

In a SGLM, the expectation $\mu(\varphi)$ of the response Y is related to the semiparametric systematic part $X^\top\vartheta+\varrho(Z)$ such that $\mu(\varphi) = h(X^\top\vartheta+\varrho(Z))$ through link h. Assuming the existence of the inverse of the mean μ, we can write $\varphi = a(X^\top\vartheta+\varrho(Z))$ with $a = \mu^{-1}\circ h$. Let $\mathcal{M}$ be a subset of ^{m} such that $X^\top\vartheta+\varrho(Z) \in \mathcal{M}$ for all $\vartheta \in \Theta$, $(X, Z) \in [c, d]^{p\times m} \times [0, 1]$ and $\varrho \in \Upsilon$. Then $a : \mathcal{M} \subset {}^m \to {}^l$ and $\Phi = a(\mathcal{M})$. We assume henceforth that Φ is a convex subset of the interior of the *natural parameter space* consisting of all φ having the finite normalizing function

$$\exp(b(\varphi)) = \int \exp(\varphi^\top y)\, d\lambda_1(y).$$

Hence in Φ, all moments of Y are finite and all the derivatives of $b(\varphi)$ exist and, in particular, the l-dimensional expectation vector and $l \times l$ covariance matrix of Y are $\mu(\varphi) = {}_{\varphi}Y = \dot{b}(\varphi) \equiv \partial b(\varphi)/\partial\varphi$ and $\Sigma(\varphi) = \mathrm{Cov}_\varphi(Y) = \ddot{b}(\varphi) \equiv \partial^2 b(\varphi)/\partial\varphi\,\partial\varphi^\top$.

The following assumption ensures that the composite a is adequately smooth for the ongoing theoretic analysis. Write $|A|$ for the Euclidean norm of matrix $A = (a_{i,j}) \in {}^{p\times m}$ so that $|A|^2 = \sum_{i=1}^p \sum_{j=1}^m a_{i,j}^2$. Let

$$s(Y, \varphi) = \frac{\partial \log f}{\partial \varphi}(y|\varphi), \quad \dot{s}(Y, \varphi) = \frac{\partial s}{\partial \varphi^\top}(Y, \varphi),$$

Assumption 2.1 The link $h : \mathcal{M} \to {}^l$ and the inverse $\mu^{-1} : h(\mathcal{M}) \to R^l$ are twice continuously differentiable. Moreover, both $s(Y, \varphi)$ and $\dot{s}(Y, \varphi)$ are $F \times G$ square-integrable for every $\vartheta \in \Theta$ and $\varrho \in \Upsilon$.

Under Assumption 2.1, a is twice continuously differentiable, so that the conditional score $s(Y, \varphi)$ and the conditional information matrix $I(\varphi)$ given X, Z can be calculated as

$$s(Y, \varphi) = \dot{a}^\top(\varphi)\big(Y - h(\varphi)\big), \quad I(\varphi) = \mathrm{Cov}_\vartheta\left(s(Y, \varphi)|X, Z\right) = \dot{a}^\top(\varphi)\Sigma(\varphi)\dot{a}(\varphi).$$

Clearly under Assumption 2.1, $I(a(X^\top\vartheta + \varrho(Z)))$ is G-integrable for $\vartheta \in \Theta$ and $\varrho \in \Upsilon$. Using the chain rule, we calculate

$$\dot{s}(Y, \varphi) = \frac{\partial s(Y, \varphi)}{\partial \varphi^\top} = \sum_{i=1}^{l} \ddot{a}_i^\top(\varphi)(y_i - h_i(\varphi)) - I(\varphi), \tag{2}$$

where y_i, a_i and h_i are the i-th components of Y, a and h respectively.

Let $\{q, q_1, q_2, \cdots\}$ be the densities of probability measures $\{Q, Q_1, Q_2, \cdots\}$ with respect to some dominating measure λ defined on the same σ-field. We say $\{Q_n\}$ has *tangent* τ with respect to Q if $\tau \in L_{2,0}(Q)$ such that

$$\int \left(\sqrt{n}(\sqrt{q_n} - \sqrt{q}) - (1/2)\tau\sqrt{q}\right)^2 d\lambda \to 0, \quad n \to \infty.$$

Let g be the density of G with respect to a σ-finite measure λ_2. Then the distribution of $\xi = (X, Z, Y)$ under the probability measure $P_{\vartheta,G,\varrho}$ has a density w.r.t. the dominating measure $\lambda_1 \times \lambda_2$, which is

$$f(y|\varphi)g(x, z) = \exp(\varphi^\top y - b(\varphi))g(x, z), \quad (y, x, z) \in {}^{l} \times [c, d]^{p\times m} \times [0, 1], \ \varphi \in \Phi.$$

Denote by G_Z the marginal distribution of Z. Write $\mathbf{1}$ the column vector of all components 1 and $\|\psi\| = \sup_{z\in[0,1]} |\psi(z)|$ the supremum norm for $\psi : [0, 1] \to R^l$. In an analogue to Lemma 3.4 of Schick (1993), one could verify the following lemma.

Lemma 2.2: *Suppose Assumption 2.1 holds. Let $\{G_n\} \subset \mathcal{G}$ be a sequence of distributions which have tangent u with respect to G. Let $\{\rho_n\} \subset C^m([0, 1])$ be a sequence such that $\int \|\sqrt{n}(\rho_n - \rho) - v\|^2 \, dG_Z = o(1)$ for some $v \in L_2^m(G_Z)$. Then $P_{\theta+\mathbf{1}/\sqrt{n},\gamma_n}$ (with $\gamma_n = (G_n, \rho_n)$) has the tangent*

$$\tau(\xi; u, v) = \big(Y - h(\eta)\big)^\top \dot{a}(\eta)(X^\top\mathbf{1} + v(Z)) + u(X, Z) \tag{3}$$

with respect to $P_{\theta,\gamma}$.

The tangent space $\mathcal{T}$ of the SGLM is the closed linear subset of $L_{2,0}(P_{\theta,\gamma})$ spanned componentwise by all the expressions (3) for $u \in L_{2,0}(G)$ and $v \in L_2^m(G_Z)$. Let us write

$$\dot{\kappa}_1 = X\dot{a}^\top(\eta)(Y - h(\eta)), \quad \dot{\kappa}_2 = u(X, Z), \quad \dot{\kappa}_3 = v^\top(Z)\dot{a}^\top(\eta)(Y - h(\eta)).$$

Let $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{T}_3$ be the closed linear subsets of $L_{2,0}(P_{\theta,\gamma})$, $L_{2,0}(G)$, and $L_{2,0}(P_{\theta,\gamma})$ spanned componentwise by $\dot{\kappa}_1$, $\dot{\kappa}_2$, $\dot{\kappa}_3$ for $u \in L_{2,0}(G)$ and $v \in L_2^l(G_Z)$ respectively. Then it follows from (3) that $\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3$. Note that $\mathcal{T}_1$ and $\mathcal{T}_2$, $\mathcal{T}_2$ and $\mathcal{T}_3$ are two orthogonal pairs. This follows from the equalities:

$$(\dot{\kappa}_1\dot{\kappa}_2) = 0, \qquad (\dot{\kappa}_2\dot{\kappa}_3) = 0, \quad u \in L_{2,0}(G), v \in L_2^m(G_Z),$$

which are readily verified by conditioning on (X, Z). The sum space $\mathcal{T}_2 + \mathcal{T}_3$ is closed, so that the projection $\dot{\kappa}_{1,*}$ of $\dot{\kappa}_1$ into it exists by the projection theorem. Then $L^*_{\theta,\gamma} = \dot{\kappa}_1 - \dot{\kappa}_{1,*}$ is the *efficient score* function for estimating θ in the presence of nuisance parameter $\gamma = (G, \rho)$. See e.g. Bickel, *et al.* (pages 70-71, 1993), or Van der Vaart (page 369, 1998). Because of the above orthogonality, $\dot{\kappa}_{1,*}$ is simply the projection of $\dot{\kappa}_1$ into $\mathcal{T}_3$, which is determined by the equations:

$$\big(\dot{\kappa}_1 \dot{\kappa}_3\big) = \quad \big(\dot{\kappa}_{1,*} \dot{\kappa}_3\big), \quad v \in L_2^m(G_Z).$$

Since each component of $\dot{\kappa}_{1,*}$ is an element of $\mathcal{T}_3$, we write

$$\dot{\kappa}_{1,*} = m_{\theta,\gamma}(Z) \dot{a}^\top(\eta) \big(Y - h(\eta)\big)$$

for some $p \times m$ matrix $m_{\theta,\gamma}(Z)$ (of which each row is a vector in $L_2^m(G_Z)$) satisfies:

$$\big(X \dot{a}^\top(\eta) \Sigma(\phi) \dot{a}(\eta) v(Z)\big) = \quad \big(m_{\theta,\gamma}(Z) \dot{a}^\top(\eta) \Sigma(\phi) \dot{a}(\eta) v(Z)\big), \quad v \in L_2^l(G_Z),$$

where we used the argument of conditioning on X, Z. Now further conditioning on Z, we obtain

$$m_{\theta,\gamma}(Z) = \quad \big(X I(\eta) | Z\big) \quad ^{-1} \big(I(\eta) | Z\big), \tag{4}$$

where $\quad ^{-1} \big(I(\eta)|Z\big)$ denotes the inverse (assuming the existence) of the matrix $\quad \big(I(\eta)|Z\big)$. Hence the efficient score function is given by

$$L^*_{\theta,\gamma}(\xi) = (X - m_{\theta,\gamma}(Z)) \dot{a}^\top(\eta) \big(Y - h(\eta)\big). \tag{5}$$

The *efficient information matrix* is

$$J_{\theta,\gamma} = \quad L^*_{\theta,\gamma}(\xi) L^{*\top}_{\theta,\gamma}(\xi) = \quad \big((X - m_{\theta,\gamma}(Z)) I(\eta) (X - m_{\theta,\gamma}(Z))^\top\big). \tag{6}$$

Let $L^\#_{\theta,\gamma}(\xi) = J^{-1}_{\theta,\gamma} L^*_{\theta,\gamma}(\xi) = J^{-1}_{\theta,\gamma}(X - m_{\theta,\gamma}(Z)) \dot{a}^\top(\eta) \big(Y - h(\eta)\big)$, where $J_{\vartheta,(G,\varrho)}$ is assumed to be positive definite for every $\vartheta \in \Theta$ and $(G, \varrho) \in \Gamma = \mathcal{G} \times \Upsilon$. Note the functional $\psi(P_{\vartheta,\gamma}) = \vartheta$ is differentiable with the derivative being the identity on ^{p}. It is readily verified

$$\int L^\#_{\theta,\gamma}(\xi) \tau(\xi; u, v)\, dP_{\theta,\gamma} = 1, \quad u \in L_{2,0}(G), \quad v \in L_2^m(G_Z).$$

Thus, it follows from the convolution theorems (see e.g. Schick(pages 1490-1491, 1993)) that $L^\#_{\theta,\gamma}(\xi)$ is the *efficient influence* function for estimating θ in the presence of nuisance parameter γ.

Example 2.3 Severini and Wong (1994) investigated estimation of quasilikelihood SGLMs with the quasilikelihood function Q prescribed by

$$Q(\mu, y) = \int_\mu^y V(s)^{-1}(s - y)\, ds,$$

where $V(\mu)$ is the variance function (of the mean μ). Let $r = \rho(z; \vartheta)$ be the solution to the equation:

$$\left(\frac{\partial}{\partial r}Q\left(h(X^\top\vartheta+r);Y\right)\Big|Z=z\right)=0. \tag{7}$$

Substituting $r = \rho(z;\vartheta)$ in the above equation, implicitly differentiating it with respect to ϑ at θ and using the structural relation (1), one finds that $\rho(z;\theta) = \rho(z)$ and that "the least-favorable curve" is $\frac{\partial}{\partial\theta}\rho(\cdot;\theta) = -m_{\theta,\gamma}(\cdot)$, provided that the quasilikelihood function is a true likelihood function; where (7) boils down to the equation:

$$\left(\dot{a}^\top(X^\top\vartheta+r))(Y-h(X^\top\vartheta+r))|Z=z\right)=0, \quad \vartheta\in\Theta, z\in[0,1] \tag{8}$$

It is readily verified that the second derivative $\ddot{\rho}(\cdot,\vartheta)$ w.r.t. ϑ exists and is continuous for $\vartheta \in \Theta$ using the nice analytic behavior of the exponential family. This fact is used in concluding Theorem 4.1 below without further referring to it.

Example 2.4 Let h be the identity link and Y have normal distribution $\mathcal{N}(\mu,\sigma_0^2)$ with known dispersion parameter $\sigma_0^2 > 0$. Then SGLM (1) simplifies to the well known PLM. In this case, a is the identity map and the efficient score (5) and efficient information (6) specialize to the popular formulae of the PLM:

$$L^*_{\theta,\gamma}(\xi)=(X-\quad(X|Z))(Y-X^\top\theta-\rho(Z)), \quad J_{\theta,\gamma}=\sigma_0^2\quad(X-\quad(X|Z))^{\otimes 2},$$

where $M^{\otimes 2}$ is the product $MM^\top$ of square matrix M. Here the efficient influence function is calculated without the independence between the error and the covariates. The independence is usually assumed in PLM.

An estimate T_n of parameter $\nu \in \Theta$ is a measurable function from $\ ^n$ into R^p. We call T_n *regular* at ν if there is a probability distribution L_0 such that

$$\mathcal{L}\big(\sqrt{n}(T_n-\nu_n)|Q_{\nu_n,\tau}\big)\Rightarrow L_0, \quad \tau\in\mathcal{T}$$

for every local sequence $\{\nu_n\} \subset \Theta$ of ν, where $Q_{\nu_n,\tau}$ denotes a smooth submodel of the model $\{P_{\vartheta,G,\varrho} : \vartheta \in \Theta,\ (G,\ \varrho) \in \Gamma\}$ with arbitrary tangent $\tau \in \mathcal{T}$. Here the left-hand side expression is the distribution of $\sqrt{n}(T_n - \nu_n)$ calculated under the assumption that $\xi_1, \xi_2, \cdots, \xi_n$ are independent with the same distribution $Q_{\nu_n,\tau}$. It follows from the convolution theorems (see e.g. Bickel,*et al.* , pages 57-73, 1993) that L_0 is the convolution of the normal distribution $\mathcal{N}(0, J^{-1}_{\nu,\gamma})$ and some other distribution M independent of the normal distribution. Namely, $L_0 = \mathcal{N}(0, J^{-1}_{\nu,\gamma}) \otimes M$. An estimate T_n of ν is regular with limiting distribution $L_0 = \mathcal{N}(0, J^{-1}_{\nu,\gamma})$ if and only if T_n satisfies the expansion

$$T_n=\nu+\frac{1}{n}\sum_{j=1}^n L^{\#}_{\nu,\gamma}(\xi_j)+o_{P_{\nu,\gamma}}(n^{-1/2}).$$

Summarizing our findings above, we have the following theorem.

Theorem 2.5 : *Suppose that Assumption 2.1 holds. Suppose that $J_{\theta,\gamma}$ is invertible. Then an estimate T_n of θ is efficient in the presence of nuisance parameter γ*

if and only if (with $\eta_j = X_j^\top \theta + \rho(Z_j)$)

$$T_n = \theta + \frac{1}{n}\sum_{j=1}^{n} J_{\theta,\gamma}^{-1}(X_j - m_{\theta,\gamma}(Z_j))\dot{a}^\top(\eta_j)\big(Y_j - h(\eta_j)\big) + o_{P_{\theta,\gamma}}(n^{-1/2}). \quad (9)$$

3. The Proposed Efficient Estimate

In this section, we propose the efficient estimate based on a preliminary $\sqrt{n}$-consistent estimate of the parameter and a uniform consistent estimate of the nonparametric part. Candidates for the preliminary estimate are discussed in the next section. The main theorem is given.

Fix $\vartheta \in \Theta$ and $z \in [0,1]$. Let $\hat{\rho}_n(z,\vartheta)$ be an estimate of $\rho(z)$ which solves the equation:

$$\sum_{j=1}^{n} K_{nj}(Z_j - z)\dot{a}^\top(X_j^\top \vartheta + r))(Y_j - h(X_j^\top \vartheta + r)) = 0, \quad (10)$$

where $K_{nj}(\cdot) = K((Z_j - \cdot)/h_n)$ is a probability kernel with bandwidth $h_n > 0$, which is chosen to be bounded, Lipschitz, and have support $[-1,1]$. It is shown below that $\hat{\rho}_n(z,\vartheta)$ is differentiable with respect to ϑ with total derivative $\dot{\hat{\rho}}_n(z,\vartheta) = \frac{\partial}{\partial\vartheta}\hat{\rho}_n(z,\vartheta)$. Let $\tilde{\theta}_n$ be a preliminary estimate of θ. Then at our disposal we have the known quantities $\hat{\eta}_j = X_j^\top \tilde{\theta}_n + \hat{\rho}_n(Z_j, \tilde{\theta}_n)$, $j = 1, \cdots, n$. We estimate the information matrix $J_{\theta,\gamma}$ by the empirical version estimate $\hat{J}_{\tilde{\theta}_n}$:

$$\hat{J}_{\tilde{\theta}_n} = (1/n)\sum_{j=1}^{n}(X_j - \hat{m}_j)I(\hat{\eta}_j)(X_j - \hat{m}_j)^\top,$$

where $\hat{m}_j = \hat{m}(Z_j, \tilde{\theta}_n)$ with $\hat{m}(\cdot,\vartheta) = -\dot{\hat{\rho}}_n(z,\vartheta)$ being an estimate of $m_{\theta,\gamma}(\cdot)$. The construction of $\hat{m}(\cdot,\theta)$ is given below. Mimicking (9), we propose an efficient estimate of the parameter θ given by

$$\hat{\theta}_n = \tilde{\theta}_n + (1/n)\sum_{j=1}^{n} \hat{J}_{\tilde{\theta}_n}^{-1}(X_j - \hat{m}_j)\dot{a}^\top(\hat{\eta}_j)\left(Y_j - h(\hat{\eta}_j)\right). \quad (11)$$

It is clear that this estimate can also be viewed as the estimate resulted from the popular *scoring method* to improve efficiency. To establish the efficiency of the proposed estimate $\hat{\theta}_n$ in (11), we need to show that $\hat{\theta}_n$ satisfies the expansion (9). For $t \in {}^p$, denote $\bar{\theta}_{nt} = \theta + n^{-1/2}t$. We require the existence of $\sqrt{n}$-consistent estimate of the parameter and an uniform consistent estimate of $\rho(z)$ and $m_{\theta,\gamma}(z)$. Write $\hat{\rho}_\vartheta(z) = \hat{\rho}_n(z,\vartheta)$ and so forth.

Assumption 3.1 There exists a sequence of $\sqrt{n}$-consistent estimates $\tilde{\theta}_n$ of θ.

Assumption 3.2 There exists a map $(\vartheta, z) \mapsto \rho_\vartheta(z), \vartheta \in \Theta, z \in [0,1]$ which is twice continuously differentiable with respect to ϑ with bounded first and second derivatives $(\vartheta, z) \mapsto \dot{\rho}_\vartheta(z)$ and $(\vartheta, z) \mapsto \ddot{\rho}_\vartheta(z)$ such that $\rho_\theta = \rho$, $\dot{\rho}_\vartheta = -m_{\vartheta,\gamma}$, and

$$\sup_{|t|\le M} \|\hat{\rho}_{\bar{\theta}_{nt}} - \rho_{\bar{\theta}_{nt}}\| = o_P(1), \quad (12)$$

$$\sup_{|t|\le M} \|\hat\rho_{\bar\theta_{nt}} - \dot\rho_{\bar\theta_{nt}}\| = o_P(1) \tag{13}$$

for every $M > 0$.

Assumption 3.3 There exist a compact neighborhood $N \subset \Theta$ of θ such that

$$\sup_{\vartheta\in N}\sup_{\varrho\in\Upsilon}(|s(Y, X^\top\vartheta + \varrho(Z))|^2 + |\dot s(Y, X^\top\vartheta + \varrho(Z))|^2) < \infty.$$

It follows from (2) that Condition 3.4 next is sufficient for Assumption 3.3.

Condition 3.4 The link $h : \mathcal{M} \to {}^l$ and the inverse $\mu^{-1} : h(\mathcal{M}) \to R^l$ of the mean μ have bounded continuous first and second total derivatives. Moreover,

$$\int \sup_{\vartheta\in N,\varrho\in\Upsilon} |\ddot b(a(x^\top\vartheta + \varrho(z)))|^2 \, dG(x, z) < \infty.$$

Remark 1 : Under Condition 3.4, the composite a has bounded continuous first and second total derivatives on $\mathcal{M}$.

We are now ready to state the main theorem with the proof given in the last section.

Theorem 3.5 : *Suppose Assumptions 2.1–3.3 hold. Assume that $J_{\theta,\gamma}$ is invertible. Then the estimate $\hat\theta_n$ given in (11) satisfies (9):*

$$\hat\theta_n = \theta + \frac{1}{n}\sum_{j=1}^{n} J^{-1}_{\theta,\gamma}(X_j - m_{\theta,\gamma}(Z_j))\dot a^\top(\eta_j)\big(Y_j - h(\eta_j)\big) + o_P(n^{-1/2}). \tag{14}$$

Hence $\hat\theta_n$ is efficient and asymptotically normal $\sqrt{n}(\hat\theta_n - \theta) \Rightarrow \mathcal{N}(0, J^{-1}_{\theta,\gamma})$.

4. Convergence of $\hat\rho_n$ and Preliminary Estimate $\tilde\theta_n$.

In this section, we discuss conditions that guarantee the solution $\hat\rho_n$ of (10) satisfies Assumption 3.2. We give the results in the case of $l = m = 1$ (assumed henceforth).

We note that equation (10) is the kernel smoothing score equation for (8) based on the score $\dot a(X^\top\theta + r)(Y - h(X^\top\theta + r))$ of the log likelihood $a(X^\top\theta + r)Y - b(a(X^\top\theta + r))$ for the SGLM. Examining the proofs of Wang and Peng (2004), we observe that the boundedness of the link h, the inverse μ^{-1} of the mean μ, and the existence of all the moments of the exponential family imply their moment assumptions. Based on this observation, we give the following theorem with the proof referred to their article. Let Θ be a compact subset of ^{p} and H is a compact subset in .

Theorem 4.1 : *Assume the following hold.*
(i) Assumption 2.1 and Condition 3.4 are satisfied.
(ii) The density g of (X, Z) is Lipschitz in z:

$$|g(x, z_1) - g(x, z_2)| \le L_g|z_1 - z_2|, \quad x \in [c, d]^p,\ z_1,\ z_2 \in [0, 1],$$

for some constant $L_g > 0$ independent of x. Moreover, the marginal density g_Z of Z is bounded away from zero: $\inf_{z\in[0,1]} q_Z(z) > 0$.
(iii) Denote $\chi(\vartheta, r, z) = \ \big(a(X^\top\vartheta + r)Y - b(a(X^\top\vartheta + r))|Z = z\big)$ and $\ddot\chi(\vartheta, r, z) =$

$\frac{\partial^2 \chi}{\partial r^2}(\vartheta, r, z)$. *Then* $\inf_{\vartheta\in\Theta}\inf_{r\in H}\inf_{z\in[0,1]}|\ddot{\chi}(\vartheta, r, z)| > 0$.
(iv) $\sup_\vartheta \sup_{r\in H}\int y^2 f(y|a(x^\top\vartheta + s))\,d\nu(y)dG(x,z) < \infty$.
(v) Denote $S_\vartheta(r, Y, X) = (Y - h(X^\top\vartheta + r))\dot{a}^\top(X^\top\vartheta + r)$. *Then* $r \mapsto S_\vartheta(r, y, x)$ *is monotone for every* $\vartheta \in \Theta, x \in [c,d]^m, y \in$. *Further,* $\rho_\vartheta(z)$ *satisfies*

$$[S_\vartheta(\rho_\vartheta(z), Y, X)|Z = z] = 0, \quad \vartheta \in \Theta, z \in [0,1],$$

and $\inf_{\vartheta\in\Theta}\inf_{|r|\le r_0}\inf_{z\in[0,1]}|D_\vartheta(r,z)| > 0$ *for some* $r_0 > 0$, *where* $D_\vartheta(r,z) = \left[\frac{\partial S_\vartheta}{\partial r}(\rho_\vartheta(z) + r, Y, X)|Z = z\right]$.
(vi) For some constant $L_\rho > 0$ *independent of* ϑ,

$$|\rho_\vartheta(z_1) - \rho_\vartheta(z_2)| \le L_\rho|z_1 - z_2|, \quad z_1, z_2 \in [0,1], \vartheta \in \Theta.$$

Then for any even integer $d \ge 2$ *with* $h_n = n^{-d/(4d+12)}$, *we have*

$$\sup_{\vartheta\in\Theta}\|\hat{\rho}_n(\cdot,\vartheta) - \rho(\cdot,\vartheta)\| = O_P(n^{-d/(4d+12)}), \quad \sup_{\vartheta\in\Theta}\|\dot{\hat{\rho}}(\cdot,\vartheta) - \dot{\rho}(\cdot,\vartheta)\| = O_P(n^{-d/(4d+12)}).$$

Hence $\hat{\rho}_n$ *fulfills Assumption 3.2.*

Preliminary Estimate $\tilde{\theta}_n$. Here we give two candidates for the preliminary estimates $\{\tilde{\theta}_n\}$ which fulfills Assumption 3.1. The moment-type estimate $\hat{\theta}_M$ of Wang and Peng (2004) is based on the following moment equations:

$$(Y^k) = \left(\exp(-b(\phi))\frac{\partial^k}{\partial\phi^k}\exp(b(\phi))\right), \quad k = 1, 2, \cdots. \tag{15}$$

Specifically, $\hat{\theta}_M$ is the solution to the following p equations.

$$\frac{1}{n}\sum_{i=1}^n Y_i = \frac{1}{n}\sum_{i=1}^n \dot{b}(\phi_{n,i}(\theta)), \tag{16}$$

$$\frac{1}{n}\sum_{i=1}^n Y_i^2 = \frac{1}{n}\sum_{i=1}^n \left(\dot{b}^2(\phi_{n,i}(\theta) + \ddot{b}(\phi_{n,i}(\theta))\right), \tag{17}$$

$$\cdots\cdots\cdots$$

$$\frac{1}{n}\sum_{i=1}^n Y_i^p = \frac{1}{n}\sum_{i=1}^n \left(\dot{b}^p(\phi_{n,i}(\theta)) + \cdots + b^{(p)}(\phi_{n,i}(\theta))\right), \tag{18}$$

where $b^{(p)}$ denotes the p-th derivative of b, and $\phi_{n,i}(\theta) = a(X_i^\top\theta + \hat{r}_n(Z_i;\theta))$ with $\hat{r}_n$ an estimate of ρ. This $\hat{r}_n$ could be any consistent estimate of ρ. For instance, the above maximum-likelihood-type estimate $\hat{\rho}_n$, denoted by $\hat{\rho}_{ML}$ henceforth. The following moment-type estimate $\hat{\rho}_M(z)$ could also be used for $\hat{r}_n$. According to Wang and Peng (2004), $\hat{\rho}_M(z)$ is based on the conditional moment equation:

$$(Y|Z) = \left(\exp(-b(\phi))\frac{\partial}{\partial\phi}\exp(b(\phi))|Z\right) = (h(X^\top\theta + \rho(Z))|Z).$$

Replacing the above conditional expectations with the ordinary kernel estimates, they gave their estimate $\hat{\rho}_M(z,\theta)$ as a solution with respect to r to the equation:

$$\frac{1}{nh_n}\sum_{j=1}^{n} Y_j K_{nj}(z) = \frac{1}{nh_n}\sum_{j=1}^{n} h(X_j^\top\theta + r)K_{nj}(z), \quad z \in [0,1], \tag{19}$$

Moment estimators may have simple structures such as explicit formulas and thus have computational ease. Peng and Wang (2004) gave several important examples that have simple and explicit formulas for the moment estimates. This method is particularly useful when the number of parameters are low.

In what follows, we shall take $\hat{r}_n$ to be either of $\hat{\rho}_{ML}$ or $\hat{\rho}_M$ and the resulting estimates of θ are denoted by $\tilde{\theta}_{ML}$ and $\tilde{\theta}_M$ respectively. As pointed out by Wang and Peng (2004), (16) and (19) are not linearly dependent if the link h is canonical (i.e., a is the identical map). If the link is not canonical, namely, a is not an identical map, one may use the solution of (19) as $\hat{r}_n$. In the canonical link, one may replace (16) with an additional moment equation:

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^{p+1} = \frac{1}{n}\sum_{i=1}^{n}\big(\dot{b}^{p+1}(\phi_{n,i}(\theta)) + \cdots + b^{(p+1)}(\phi_{n,i}(\theta))\big). \tag{20}$$

For unambiguity let us focus on the system of equations (17)-(18) and (20) while $\hat{\rho}_\theta$ is an estimate of ρ. Other combinations of equations may be analogously considered. Let $A(y) = (y^2, \cdots, y^{m+1})^\top$ and $B(\phi)$ be a m-dimensional vector with components $B_1(\phi) = \dot{b}^2(\phi) + \ddot{b}(\phi)$, $B_2(\phi) = \dot{b}^3(\phi) + 3\dot{b}(\phi)\ddot{b}(\phi) + b'''(\phi)$, $\cdots$, and $B_m(\phi) = \dot{b}^{m+1}(\phi) + \cdots + b^{(m+1)}(\phi)$. Then (17)-(18) and (20) can be written in the vector form:

$$\Lambda_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\Big(A(Y_i) - B\big(a(X_i^\top\theta + \hat{\rho}_\theta(Z_i))\big)\Big) = 0. \tag{21}$$

Clearly, the above equation may have many solutions. Henceforth denote Θ a compact set in ^{p} such that equations (21) has a unique solution in Θ. Then the above equation has only one solution almost surely for large n in Θ.

Theorem 4.2 : *(i) Suppose all assumptions in Theorem 4.1 hold.*
(ii) In addition, for $\vartheta \in \Theta$, the total derivative $\dot{B}\big(a(X^\top\vartheta + \rho_\vartheta(Z))\big) = \frac{\partial B}{\partial\vartheta^\top}\big(a(X^\top\vartheta + \rho_\vartheta(Z))\big)$ exists and is continuous, square-integrable, nonsingular; and the expectation $D(\vartheta)$ is uniformly bounded from below on Θ: $\inf_{\vartheta\in\Theta}|D(\vartheta)| > 0$.Further,

$$\sup_{|r|\le r_0}\sup_{\vartheta\in\Theta}|\dot{B}\big(a(X^\top\vartheta + \rho_\vartheta(Z)+r)\big)| < \infty, \qquad \sup_{|r|\le r_0}\sup_{\vartheta\in\Theta}|\ddot{B}\big(a(X^\top\vartheta + \rho_\vartheta(Z)+r)\big)| < \infty.$$

Then the $\tilde{\theta}_{ML}$ resulted from $\hat{r}_n = \hat{\rho}_{ML}$ fulfills Assumption 3.1.

Theorem 4.3 : *Suppose all assumptions in Theorem 4.2 hold except that (ii) is replaced with a weaker one: (ii) holds with a the identity map. Then all the results of Theorem 4.2 hold with $\tilde{\theta}_M$ replaced with $\tilde{\theta}_{ML}$ resulted from $\hat{r}_n = \hat{\rho}_M$.*

5. A Proof and a Lemma

Proof of Theorem 3.5. We first note that the desired (14) is implied by the following (22) and (23):

$$\hat{J}_{\tilde{\theta}_n} = J_\theta + o_P(1) \tag{22}$$

$$\frac{1}{n}\sum_{j=1}^{n}\Big(\hat{\psi}(\xi_j,\tilde{\theta}_n) - \psi(\xi_j,\theta)\Big) + J_\theta(\tilde{\theta}_n - \theta) = o_P(n^{-1/2}) \tag{23}$$

where $\hat{\psi}(\xi_j,\tilde{\theta}_n) = (X_j - \hat{m}_{\tilde{\theta}_n}(Z_j))\dot{a}^\top(\hat{\eta}_j)(Y_j - h(\hat{\eta}_j))$ and $\psi(\xi,\vartheta) = (X - m_{\vartheta,(G,\rho_\vartheta)}(Z))\dot{a}^\top(\zeta)(Y - h(\zeta))$ with $\zeta = X^\top\vartheta + \rho_\vartheta(Z)$ so that, in view of $\rho_\theta = \rho$, we have $\psi(\xi_j,\theta) = (X_j - m_{\theta,\gamma}(Z_j))\dot{a}^\top(\eta_j)(Y_j - h(\eta_j))$. Since $\hat{J}_{\tilde{\theta}_n}$ is the empirical version estimate of $J_{\theta,\gamma}$, it follows that (22) is implied by (23) by the continuity and boundedness of the involved functions. Thus, we only have to show (23) which, in view of Assumption 3.1, is implied by the following (24) and (25).

$$\sup_{|t|\le M}\left|\frac{1}{n}\sum_{j=1}^{n}\Big(\hat{\psi}(\xi_j,\bar{\theta}_{nt}) - \psi(\xi_j,\bar{\theta}_{nt})\Big)\right| = o_P(n^{-1/2}) \tag{24}$$

and

$$\sup_{|t|\le M}\left|\frac{1}{n}\sum_{j=1}^{n}\Big(\psi(\xi_j,\bar{\theta}_{nt}) - \psi(\xi_j,\theta) + J_\theta n^{-1/2}t\Big)\right| = o_P(n^{-1/2}) \tag{25}$$

for arbitrary $M > 0$. An application of Lemma 5.1 below with $w(\xi,\vartheta) = \psi(\xi,\vartheta)$ yields (25), using equality $M_\theta = J_{\theta,\gamma}$ which is easily verified by means of the equality $E(u(X,Z)(Y - h(\eta)) = E(u(X,Z)E(Y - h(\eta)|X,Z) = 0$ for any $u \in L_2(G)$. To show (24), we bound its left hand side by $A_n + B_n$ where

$$A_n = \sup_{|t|\le M}\left|\frac{1}{n}\sum_{j=1}^{n}\left(\hat{m}_{\bar{\theta}_{nt}}(Z_j) - m_{\bar{\theta}_{nt},\gamma}(Z_j)\right)s(Y_j,\bar{\eta}_{ntj})\right|$$

and

$$B_n = \sup_{|t|\le M}\left|\frac{1}{n}\sum_{j=1}^{n}\left(X_j - m_{\bar{\theta}_{nt},\gamma}(Z_j)\right)\left(s(Y_j,\bar{\eta}_{ntj}) - s(Y_j,\eta_{ntj})\right)\right|,$$

with $\bar{\eta}_{ntj} = X_j^\top\bar{\theta}_{nt} + \hat{\rho}_{\bar{\theta}_{nt}}(Z_j)$ and $\eta_{ntj} = X_j^\top\bar{\theta}_{nt} + \rho_{\bar{\theta}_{nt}}(Z_j)$. Note

$$\begin{aligned}
A_n &\le \sup_{|t|\le M}\left\{\left\|\hat{m}_{\bar{\theta}_{nt}} - m_{\bar{\theta}_{nt},\gamma}\right\|(1/n)\sum_{j=1}^{n}|s(Y_j,\bar{\eta}_{ntj})|\right\} \\
&\le \sup_{|t|\le M}\left\|\dot{\hat{\rho}}_{\bar{\theta}_{nt}} - \dot{\rho}_{\bar{\theta}_{nt}}\right\|\sup_{|t|\le M}\left\{(1/n)\sum_{j=1}^{n}\left(|s(Y_j,\eta_j)| + |s(Y_j,\bar{\eta}_{ntj}) - s(Y_j,\eta_j)|\right)\right\}
\end{aligned}$$

By the mean value theorem of a vector function,

$$s(Y_j,\bar{\eta}_{ntj}) - s(Y_j,\eta_j) = \int_0^1 \dot{s}^\top\big(Y_j,\eta_j + u(\bar{\eta}_{ntj} - \eta_j)\big)du\big(\hat{\rho}_{\bar{\theta}_{nt}}(Z_j) - \rho(Z_j)\big).$$

Since

$$\eta_j + u(\bar{\eta}_{ntj} - \eta_j) = X_j^\top(\theta + un^{-1/2}t) + \rho(Z_j) + u(\hat{\rho}_{\bar{\theta}_{nt}}(Z_j) - \rho(Z_j)),$$

it follows from Assumption 3.3 and (12) that $\dot{s}^\top\big(Y_j, \eta_j + u(\bar{\eta}_{ntj} - \eta_j)\big)$ is bounded by a square-integrable random variable $R(\xi_j)$ independent of u, t for sufficiently large n. Combining these, using (13) and applying the central limit theorem, we conclude $A_n = o_P(n^{-1/2})$. Analogously one may show $B_n = o_P(n^{-1/2})$. This completes the proof of the theorem. □

Let $(\mathcal{X}, P_\vartheta)$ be a probability space space indexed by $\vartheta \in \Theta \subset {}^p$. Denote $P = P_\theta$ with $\theta \in \Theta$. Let $\xi_1, \xi_2, \cdots, \xi_n$ be independent random variables with a common distribution under P.

Lemma 5.1: *Suppose there exists a neighborhood N of θ such that*
(a) For each $x \in \mathcal{X}$, the map $\vartheta \mapsto w(x, \vartheta)$ is continuously differentiable on N with gradient $\vartheta \mapsto \dot{w}(x, \vartheta)$.
(b) There is an integrable function H such that $\sup_{\vartheta \in N} |\dot{w}(x, \vartheta)| \leq H(x)$ for all $x \in \mathcal{X}$. Then for a sequence $\{r_n\}$ converging to 0,

$$\sup_{|\vartheta - \theta| \leq a_n} |\dot{W}_n(\vartheta) - M_\theta| \to 0, \quad a.s. \tag{26}$$

$$\sup_{|\vartheta - \theta| \leq a_n} \frac{|W_n(\vartheta) - W_n(\theta) - M_\theta(\vartheta - \theta)|}{|\vartheta - \theta|} \to 0, \quad a.s. \quad n \to \infty. \tag{27}$$

where

$$W_n(\vartheta) = \frac{1}{n}\sum_{j=1}^n w(\xi_j, \vartheta), \quad M_\theta = \int \dot{w}(x, \theta)\, dP(x).$$

Proof: For $r > 0$ let ω_r denote the map defined by

$$\omega_r(x) = \sup_{|\vartheta - \theta| \leq r} |\dot{w}(x, \vartheta) - \dot{w}(x, \theta)|, \quad x \in \mathcal{X}.$$

Then for each $x \in \mathcal{X}$, $\omega_r(x) \downarrow 0$ as $r \downarrow 0$; moreover, $0 \leq \omega_r(x) \leq H(x)$. Thus an application of the Lebesgue dominated convergence theorem yields

$$\lim_{r \to 0} \int \omega_r\, dP = 0.$$

In addition, for every $r > 0$,

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^n \omega_{r_n}(\xi_j) \leq \limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^n \omega_r(\xi_j) = \int \omega_r\, dP.$$

Letting $r \to 0$ gives

$$\frac{1}{n}\sum_{j=1}^n \omega_{r_n}(\xi_j) \to 0, \quad a.s. \quad n \to \infty. \tag{28}$$

Since the left hand side of (26) is bounded by

$$|\dot{W}_n(\theta) - M_\theta| + \frac{1}{n}\sum_{j=1}^{n} \omega_{r_n}(\xi_j),$$

which converges to 0 almost surely in view of (28) and the strong law of large numbers. This immediately results in (26). By the Taylor theorem, each coordinate of $|\vartheta - \theta|(W_n(\vartheta) - W_n(\theta) - M_\theta(\vartheta - \theta))$ is bounded by the left hand side of (26) provided $|\vartheta - \theta| \le r_n$. Thus (26) implies (27). This completes the proof. □

References

[1] Bickel, P.J. (1982). On adaptive estimation. *Ann. Statist.*, **10**, 647–671.

[2] Bickel, P.J., Klaassen, C.A.J., Ritov, Y., and Wellner, J.A. (1993). *Efficient and adaptive estimation in semiparametric models.* Johns Hopkins Univ. Press, Baltimore.

[3] Bhattacharya and Zhao (1993). Semiparametric inference in a partially linear model. *Ann. Statist.*, **25**, 244–262.

[4] Chen, H. (1995). Asymptotically efficient estimation in semiparametric generalized models. *Ann. Statist.*, **23**, 1102 - 1120.

[5] Enno, M. and Van der Geer, G. (1997). Penalized quasi-likelihood estimation in partial linear models. *Ann. Statist.*, **25**, 1014 - 1035.

[6] Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimate in generalized linear models. *Ann. Statist.*, **13**, 342 - 368.

[7] McCullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models.* 2nd edition. Chapman and Hall, London.

[8] Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. *J. R. Statis. Soc. A*, **135** 370 - 384.

[9] Forrester, J.S., Hooper, W. J., Peng, H., and Schick, A. (2003). On the construction of efficient estimates in semiparametric models. *Statistics and Decisions*, **21**, 109-138.

[10] Severini, T. A. and Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *Ann. Statist.*, **20**, 1768 - 1802.

[11] Severini, T. A. and Staniswalis, J. G. (1994). Quasi-likelihood Estimation in Semimparametric Models. *J. Amer. Statist. Assoc.*, **89**, 502 - 511.

[12] Schick, A. (1986). On asymptotically efficient estimation in semiparametric models. *Ann. Statist.*, **14**, 1139–1151.

[13] Schick, A. (1987). A note on the construction of asymptotically linear estimates. *J. Statist. Planning Inference*, **16**, 89-105. Correction, **22** (1989), 269–270.

[14] Schick, A. (1993). On efficient estimation in regression models. *Ann. Statist.*, **21**, 1486–1521. Correction and addendum(1995), **23**, 1862–1863.

[15] Schick, A. (2001). Sample splitting with Markov chains. *Bernoulli*, **7**, 243–266.

[16] Van der Vaart, A.W. (1998). Asymptotic Statistics. *Cambridge University Press.*

[17] Wang, X. and Peng, H. (2004). Moment estimation in a semiparametric generalized linear model. To appear in *Statist. & Probabil. Lett.*