$M-$ESTIMATION INFERENCE FOR PARTIALLY LINEAR SINGLE-INDEX MODELS: AN EMPIRICAL LIKELIHOOD APPROACH

Matthieu Marbac [1] & Valentin Patilea [2]

[1] *CREST, Ensai, Campus de Ker-Lann, 35172 Bruz, France; matthieu.marbac-lourdelle@ensai.fr*
[2] *CREST, Ensai, Campus de Ker-Lann, 35172 Bruz, France; valentin.patilea@ensai.fr*

Résumé. Les modèles partiellement linéaires à un indice sont des outils utiles pour capturer les relations entre une variable réponse et un ensemble de covariables potentiellement grand. L'approximation de la réponse est donnée par la somme d'un terme linéaire et d'une fonction de lien non-paramétrique appliquée à une seconde combinaison linéaire de covariables, généralement appelée l'indice. Cette approximation est définie par rapport à une fonction de perte qui caractérise un modèle de loi pour la variable réponse conditionnellement aux covariables. Nous considérons une famille générale de fonctions de perte et étudions le modèle de régression partiellement linéaire à un indice correspondant. Mis à part que certains moments doivent être finis, la loi conditionnelle du terme d'erreur peut être aussi générale que possible. L'estimation se fait par vraisemblance empirique via une condition sur les moments dans laquelle nous utilisons un estimateur de la fonction de lien. Nous montrons la pivotalité asymptotique du rapport de vraisemblance sous des conditions peu restrictives. Nous proposons une procédure automatique simple qui permet de régler les paramètres nécessaires à l'estimation de la fonction de lien.

Mots-clés. Lissage à noyaux, modèles semi-paramétriques, moments conditionnels, rééchantillonnage multiple, théorème de Wilks.

Abstract. Partially linear single-index models represent a versatile tool to capture the relationship between response variables and possibly high-dimensional covariate vectors. The approximation of the response is given by the sum of a linear term and of a nonparametric link function of a second linear combination of covariates, usually called the index. This approximation is defined with respect to a loss function which characterizes a feature of the conditional law of the response given the covariates. We consider a general family of loss functions and investigate the corresponding partially linear single-index regression models. Except for imposing some moments to be finite, the conditional law of the error term is allowed to be general. For the inference, we adopt the empirical likelihood (EL) approach based on a class of moment conditions in which we plug-in estimates of the nuisance link function. We show the asymptotic pivotality of the likelihood ratio under weak high-level conditions. A simple data-driven choice of the tuning parameter for the estimation of the link function is proposed.

Keywords. conditional moments, kernel smoothing, multiplier bootstrap, semiparametric models, Wilks Theorem

1 Introduction

The observations are realizations of some random covariate vectors $X \in \mathbb{R}^{d_X}$ and $W \in \mathbb{R}^{d_W}$ and of a response variable $Y \in \mathbb{R}$. The components of the vectors X and W could be continuous or discrete random variables. Consider a loss function $\mathcal{L}(u;v) = L(u-v)$, $u, v \in \mathbb{R}$, with $L(\cdot)$ a nonnegative piecewise differentiable convex function such that $L(0) = 0$, and let $\xi(\cdot)$ be its piecewise derivative. For instance, we can consider the quadratic loss ($L(v) = v^2$, $\xi(v) = 2v$), the quantile loss (for $\tau \in (0,1)$, $L(v) = |v| + (2\tau - 1)v$, $\xi(v) = 2(\tau - \mathbf{1}\{v \leq 0\})$), the expectile loss (for $\tau \in (0,1)$, the loss is $L(v) = |\tau - \mathbf{1}\{v \leq 0\}|v^2$ and $\xi(v) = 2v\{(1-\tau)\mathbf{1}\{v \leq 0\} + \tau\mathbf{1}\{v > 0\}\}$)...

We propose a general semiparametric partially linear single-index model (PLSIM)

$$Y = X^\top\theta_1 + h(W^\top\theta_2) + \varepsilon, \qquad \text{with} \qquad \mathbb{E}(\xi(\varepsilon) \mid X, W) = 0 \quad \text{a.s.}, \tag{1}$$

where $\theta = (\theta_1^\top, \theta_2^\top)^\top \in \Theta = \Theta_1 \times \Theta_2 \subset \mathbb{R}^{d_\theta}$, $d_\theta = d_X + d_W$, are unknown parameters and $h(\cdot)$ is an unknown univariate real-valued function (often seen as a nuisance function). Let $\theta_0 = (\theta_{0,1}^\top, \theta_{0,2}^\top)^\top$ be the true value of θ and $h_0(\cdot)$ the true function of $W^\top\theta_{0,2}$ in the model (1). The error term ε is not necessarily independent of X and W, we only impose the identification condition that $\xi(\varepsilon)$ has a zero conditional expectation given the covariates. The nonparametric part of the model represented by the function h_0 could absorb any constant in its value and in its argument. One common approach to restrict θ_2, for identification purposes, is to set the norm of θ_2 equal to 1 and the sign of one of its components.

Our framework includes many semi-parametric models. The single-index models represent the particular case where the linear part $X^\top\theta_1$ no longer appears ($d_\theta = d_W$). Kong & Xia (2012), Ma & He (2016), Zhao *et al.* (2017) considered the single-index modeling in the quantile regression context, which could be obtained here with the quantile loss. Xue & Zhu (2006) studied the single-index mean regression, a context that we obtain with the quadratic loss. Zhu & Xue (2006) investigated the partially linear single-index mean regression model. If W is a real-valued covariate, we recover the generalized partially linear model (here $d_\theta = d_X + 1$). See Robinson (1988) for the case of the mean regression. See Boente *et al.* (2006) for the case of a robust regression. However, our framework covers a much larger set of interesting situations that has not been yet investigated in the literature, such as the partially linear single-index quantile, robust or expectile regressions. Moreover, the model (1) allows other conditional moments to have an unknown form. For instance, we allow for conditional variance of unknown form.

This paper is organized as follows. Section 2 presents the equivalent moment conditions and Section 3 details the inference of the new approach. Numerical experiments are conducted in Section 4 and a conclusion is given in Section 5.

2 Equivalent moment conditions

The general model introduced in equation (1) could be rewritten under the form of a conditional moment equation

$$\mathbb{E}(\rho(Z;\theta,h) \mid X,W) = 0 \quad \text{a.s.}, \tag{2}$$

where $Z = (Y, X^\top, W^\top)^\top$ and $\rho(Z;\theta,h) = \xi\left(Y - X^\top\theta_1 - h(W^\top\theta_2)\right) \in \mathbb{R}$. The model (2) requires a methodology for estimating θ and h, with h in a function space. A common approach to avoid a simultaneous search involving an infinite-dimensional parameter is the profiling (see Severini & Wong (1992), Liand *et al.* (2010) and Zhang *et al.* (2017) for the profiling approach in the PLSIM context for mean and quantile regression, respectively). Here, for each w in the support of W, and any $\theta = (\theta_1^\top, \theta_2^\top)^\top$, let

$$h_\theta(t) = \arg\min_a \mathbb{E}(L(Y - X^\top\theta_1 - a) \mid W^\top\theta_2 = t), \qquad t \in \mathbb{R}. \tag{3}$$

If $L(\cdot)$ is strictly convex, the function $h_\theta(\cdot)$ is uniquely defined. As usually in PLSIM, it will be assumed that $h_{\theta_0}(W^\top\theta_{0,2}) = h_0(W^\top\theta_{0,2})$. Hence, one expects that, for each x, w, the value θ_0 realizes the minimum of $\theta \mapsto \mathbb{E}(L(Y - X^\top\theta_1 - h_\theta(W^\top\theta_2)) \mid X = x; W = w)$. To proceed towards the empirical likelihood inference, we use the following lemma which shows that (2) can be written as a unconditional moment condition.

Lemma 1 *Let $\omega(X,W)$ some positive weight function of X and W. Under mild conditions, assume that the model identification condition* (2) *holds true, then there exists a neighborhood of (θ_0, h_{θ_0}) over wich*

$$\mathbb{E}(\rho(Z;\theta_0,h_{\theta_0})\mathbf{J}(\theta_0)^\top\nabla_\theta\mathbb{E}(\rho(Z;\theta_0,h_0) \mid X,W)\omega(X,W)) = 0 \Leftrightarrow (\theta,h) = (\theta_0,h_{\theta_0}) \tag{4}$$

where $\mathbf{J}(\theta_0)$ is a Jacobian matrix implied by the identifiability constraints made on θ and $\omega(X,W)$ is any positive weight function of X and W.

The weight function $\omega(\cdot)$ could help to obtain the pivotalness, to avoid handling denominators that could be close to zero, to avoid estimating unknown positive functions, and to improve the quality of the inference. These properties motivate the moment condition we will propose in the following for empirical likelihood inference.

3 Plug-in empirical likelihood for PLSIM

Let $Z_1, \ldots, Z_n$ be a random sample of $Z = (Y, X^\top, W^\top)^\top$. When the true function $h_0(\cdot)$ is known, hypothesis testing $\theta = \theta_0$ can be done by using the empirical likelihood ratio

$$\mathrm{EL}_n(\theta, h_0) = \max\left\{\sum_{i=1}^n \ln(np_i) : p_i > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i m_i(\theta, h_0) = 0\right\}. \tag{5}$$

where $m_i(\theta, h) = \rho(Z_i; \theta, h)\mathbf{J}(\theta)^\top \nabla_\theta \mathbb{E}(\rho(Z_i; \theta, h) \mid X_i, W_i)\omega(X_i, W_i)$. Using the results presented in Owen (1990), Qin & Lawless (1994) or Owen (2001), it can shown that

$$-2 \log \mathrm{EL}_n(\theta_0, h_0) \to_d \chi^2_{d_\theta - 1}, \tag{6}$$

where χ^2_p denotes a chi-square random variable with p degrees of freedom. This allows to build a confidence region $\{\theta : \mathrm{EL}_n(\theta, h_0) > c\}$ for θ_0, where c is a suitable threshold.

We show that, under some assumptions, the estimation of the nuisance function h is negligible for hypothesis testing, when a suitable weight function is considered.

Theorem 1 (Wilks' Theorem) *In the class of general PLSIM introduced in equation (1), using the moment equations we introduced in the previous section, with a* suitable weight function $\omega(X, W)$*, the pivotality property (6) holds when* h_0 *is replaced by suitable estimators* $\widehat{h}$*, that is*

$$-2 \log \mathrm{EL}_n(\theta_0, \widehat{h}) \to_d \chi^2_{d_\theta - 1}.$$

With the suitable choice of the weight function $\omega(X, W)$, the high-level conditions we impose on $\widehat{h}$ are mainly of two types: on one hand to belong to a suitable Donsker class of smooth functions with probability tending to 1, and, on the other hand, to converge uniformly to h_0 at the rate $o_{\mathbb{P}}(n^{-1/4})$. Common estimators, such as kernel-based estimators, satisfy our high-level conditions.

Nonparametric estimates $\widehat{h}$ require a rule for the smoothing parameter. Our extensive empirical investigation shows that results are sensitive to this rule. The pivotality of the empirical log-likelihood allows us to propose a novel data-driven rule for selecting the smoothing parameter. Our rule is based on multiplier bootstrap, is very easy to implement and perform quite well in applications.

4 Some empirical evidence

Data are simulated as follows. We consider the following independent variables $A \sim \mathcal{B}e(2, 2)$, $B \sim \mathcal{B}e(2, 2)$, $C_j \sim \mathcal{B}e(2, 2)$ for $j = 1, \ldots, 5$. The dimensions are $d_X = 1$, $d_W = 4$, the observed covariates are defined by $W_j = A + B + C_j$ for $j = \{1, \ldots, 4\}$, $X = A + C_5$, and the parameters are $\theta_{0,1} = 1$ and $\theta_{0,2} = \begin{bmatrix} 1 & -1 & 1 & -1 \end{bmatrix}^\top$. We consider two cases of PLSIM: a 25%-quantile regression with heteroscedastic shifted exponential noise and a 25%-expectile regression with a heteroscedastic Gaussian noise. Two samples sizes are considered (100 and 500), and for each situation 5000 replicates are generated. The estimation of h is achieved with the locally linear method by considering bandwidths $b = n^{-1/3.1} + (n^{-1/6.9} - n^{-1/3.1})k/20$ for $k = 0, \ldots, 20$. Bandwidth selection is done with $R = 10^3$ multiplier bootstrap samples.

We want to illustrate the behavior of the empirical likelihood ratio test and the fact that the estimation of the non-parametric part is negligible. Thus, we test different values for the parameter: the true parameters and some shifted values (*i.e.,* we test $\theta_0 = \tilde{\theta}$ where $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2)$, $\tilde{\theta}_1 = \theta_{0,1} + c$ and $\tilde{\theta}_2 = \theta_{0,2}\sqrt{1-(c/10)^2} + c|\theta_{0,2}|/10$ for some deviation c). Table 1 presents the empirical probabilities of rejection where the nominal level is 0.05. We see that without deviation (*i.e.,* the true parameter is tested), the nominal level is reached when the sample size is large enough. Moreover, the probability of rejection tends to one when the sample size increases for any deviation. Finally, note that the impact of the estimation of function $h(\cdot)$ vanishes when n increases.

	Quantile regression				Expectile regression			
deviation c	$n = 100$		$n = 500$		$n = 100$		$n = 500$	
	h known	h estimated	h known	h estimated	h known	h estimated	h known	h estimated
-1.6	0.65	0.88	1.00	1.00	0.56	0.48	1.00	0.96
-1.2	0.65	0.67	1.00	1.00	0.43	0.34	0.95	0.79
-0.8	0.64	0.39	1.00	0.95	0.30	0.24	0.69	0.45
-0.4	0.39	0.18	0.85	0.37	0.18	0.18	0.24	0.16
0	0.10	0.12	0.06	0.06	0.13	0.15	0.07	0.08
0.4	0.13	0.19	0.30	0.40	0.15	0.16	0.21	0.14
0.8	0.21	0.39	0.71	0.95	0.25	0.21	0.64	0.41
1.2	0.29	0.63	0.90	1.00	0.38	0.31	0.92	0.77
1.6	0.37	0.81	0.97	1.00	0.52	0.45	0.99	0.95

Table 1: Empirical probabilities of rejection for testing the parameters by considering different deviations of the true parameters θ_0 and a nominal level of 0.05.

5 Conclusion

We proposed a new EL-based approach for a general class of PLSIM. This method is based on an equivalent characterization of the initial conditional moment restriction using unconditional moment restriction. The pivotality of the likelihood ratio EL remains valid despite the fact that the nuisance function is estimated. This approach could be easily extended to other situations. For instance, we could consider the case of a multi-dimensional response variable (which implies multiple conditional moment restrictions). The approach could also be extended to generalized PLSIM (*e.g.,* binary response variable modeled by a semi-parametric partially linear single index logit model).

Bibliographie

Boente, G., He, X., and Zhou, J. (2006). *Robust estimates in generalized partially linear models.* Ann. Statist., 34(6):2856–2878.

Kong, E. and Xia, Y. (2012). *A single-index quantile regression model and its estimation.* Econometric Theory, 28(4):730–768.

Liang, H., Liu, X., Li, R., and Tsai, C.-L. (2010). *Estimation and testing for partially linear single-index models.* Ann. Statist., 38(6):3811–3836.

Ma, S. and He, X. (2016). *Inference for single-index quantile regression models with profile optimization.* Ann. Statist., 44(3):1234–1268.
Owen, A. (1990). *Empirical likelihood ratio confidence regions.* Ann. Statist., 18(1):90-120
Owen, A. B. (2001). *Empirical likelihood.* Chapman and Hall/CRC.
Qin, J. and Lawless, J. (1994). *Empirical likelihood and general estimating equations.* Ann. Statist., 22(1):300–325.
Robinson, P. M. (1988). *Root-N -consistent semiparametric regression.* Econometrica, 56(4):931–954.
Severini, T. A. and Wong, W. H. (1992).*Profile likelihood and conditionally parametric models.* Ann. Statist., 20(4):1768–1802.
Xue, L.-G. and Zhu, L. (2006). *Empirical likelihood for single-index models.* Journal of Multivariate Analysis, 97(6):1295 – 1312.
Zhang, Y., Lian, H., and Yu, Y. (2017). *Estimation and variable selection for quantile partially linear single-index models.* Journal of Multivariate Analysis, 162:215 – 234.
Zhao, W., Lian, H., and Liang, H. (2017). *Quantile regression for the single-index coefficient model.* Bernoulli, 23(3):1997–2027.
Zhu, L. and Xue, L. (2006). *Empirical likelihood confidence regions in a partially linear single-index model.* J. R. Stat. Soc. Ser. B Stat. Methodol., 68(3):549–570.