Kyoto University,
Graduate School of Economics
Research Project Center Discussion Paper Series

Focused Information Criterion for Series Estimation in Partially Linear Models

Naoya Sueishi and Arihiro Yoshimura

Discussion Paper No. E-14-001

Research Project Center
Graduate School of Economics
Kyoto University
Yoshida-Hommachi, Sakyo-ku
Kyoto City, 606-8501, Japan

April, 2014

Focused Information Criterion for Series Estimation in Partially Linear Models

Naoya Sueishi* and Arihiro Yoshimura†

April 2014

Abstract

This paper proposes a focused information criterion (FIC) for variable selection in partially linear models. Our criterion is designed to select an optimal model for estimating a focus parameter, which is a parameter of interest. We estimate the model by the series method and jointly select the variables in the linear part and the series length in the nonparametric part. A Monte Carlo simulation shows that the proposed FIC successfully selects the model that has a relatively small mean squared error of the estimator for the focus parameter.

JEL Classification Numbers: C14, C52, C53.

*Graduate School of Economics, Kyoto University, Email: sueishi@econ.kyoto-u.ac.jp.

†Graduate School of Economics, Kyoto University. Research Fellow of Japan Society for the Promotion of Science. Yoshida-Hommachi, Sakyo-ku, Kyoto, Kyoto, 606-8501, Japan. Tel: +81-75-753-7197. Email: arihiroyoshimura@gmail.com

1 Introduction

The partially linear model (PLM) is one of the most popular semiparametric models in econometrics and statistics. It retains the nice interpretability of linear models and flexibility of nonparametric models while avoiding the curse of dimensionality. The first application of the PLM was reported by Engle et al. (1986), who investigated a relationship between electricity sales and temperature. Since then, many estimation and inference methods have been advocated by Chen (1988), Robinson (1988), Speckman (1988), Donald and Newey (1994), Liang et al. (1999), and Zhang et al. (2011). Härdle et al. (2013) included many recent applications of the PLM in their study.

This paper proposes a new model selection criterion for a PLM that is estimated by the series estimator of Donald and Newey (1994). We develop a criterion for jointly selecting variables in the linear part and the series length in the nonparametric part. We introduce the local misspecification framework of Claeskens and Hjort (2003) and derive a focused information criterion (FIC) that allows us to select an optimal model depending on a specific parameter of interest.

FIC has attracted increasing attention in recent model selection literature. In regression, standard selection criteria, such as the Akaike information criterion (AIC; Akaike, 1973), the Bayesian information criterion (BIC; Schwarz, 1978), and Mallows C_p (Mallows, 1973), are designed to select a model with a good fit or a good prediction property. However, rather than being interested in finding a model with a good fit, applied researchers in econometrics often focus on identifying a model that estimates a few parameters with specific economic implications. FIC is designed to select an optimal model for estimating the specific parameter of interest, which is called the *focus parameter*. Our goal is to select the model that attains the smallest mean squared error (MSE) of the estimator for the focus parameter. The FIC is derived as an approximately unbiased estimator for the MSE.

Our setup is different from that of existing FIC in that our full model is misspecified. In the FIC framework, there is a full (largest) model that nests the other candidate submodels. It is assumed that the full model is correctly specified and that the other submodels are misspecified (underspecified). Although misspecification can be avoided by using the full model, the variance of the estimator can be large when the full model contains too many parameters to estimate. FIC evaluates the trade-off between misspecification bias and estimation variance. In contrast to the standard setup, our candidate models, including the full model, are misspecified because of the nature of the series estimator. Because a finite-dimensional series is only an approximation of the true unknown function, all candidate models are intrinsically misspecified. Therefore, we construct an FIC for the pseudo-true parameter values rather than for the true parameter values. The pseudo-true values are defined as the linear projection coefficients of the full model. Our simulation result shows that our FIC works well even when the performance is evaluated at the true parameter values.

Our FIC is similar to that of Claeskens and Hjort (2003), who proposed an FIC for linear regression models with a finite number of regressors. However, they considered their full model to be the true model, whereas we consider our full model to be an approximation of the true model. Liu (2013) investigated the properties of model averaging estimators under a setting similar to that of Claeskens and Hjort (2003).

Other related studies include those by Ding et al. (2011) and Zhang and Liang (2011). Ding et al. (2011) proposed a variable selection method for the PLM on the basis of lasso. Their setting is similar to ours, but they did not address the problem of finding an optimal model for estimating the focus parameter. Zhang and Liang (2011) considered the generalized additive PLM and proposed an FIC for variable selection in the linear part of the model. Their FIC is not applicable to determining the series length, whereas our FIC can simultaneously determine the variables in the linear part and the series length.

The remainder of the paper is organized as follows. Section 2 introduces the model and describes our local misspecification framework. Section 3 derives the FIC. Section 4 presents the result of a Monte Carlo study. Section 5 presents our conclusions.

2 Model

We consider the PLM

$$y_i = x_i'\beta + g(z_i) + u_i, \quad E[u_i|x_i, z_i] = 0, \quad i = 1, \ldots, n, \tag{1}$$

where y_i is a scalar dependent variable and x_i is a finite-dimensional vector of regressors. For simplicity of notation, we assume that z_i is a scalar. The functional form of $g(z)$ is unknown. Heteroskedasticity of unknown form is allowed.

The vector β and the function $g(z)$ are estimated jointly by the series estimator of Donald and Newey (1994). The series estimator approximates $g(z)$ by a linear combination of a $J \times 1$ vector of basis functions: $p^J(z) = (p_1(z), \ldots, p_J(z))'$. The series can be a power series, piecewise local polynomial spline, Fourier series, among others. For a power series, the vector of basis functions is $p^J(z) = (1, z, z^2, \ldots, z^{J-1})'$. The estimator is obtained by regressing y_i on x_i and $p^J(z_i)$.

We first consider variable selection in the linear part of (1). Suppose that x_i is divided into $p \times 1$ and $q \times 1$ subvectors: $x_i = (x_{1,i}', x_{2,i}')'$. Also, β is divided as $\beta = (\beta_1', \beta_2')'$ so that it is conformable with x_i. The vector $x_{1,i}$ is the set of regressors that must be included in the model. For instance, $x_{1,i}$ may be a regressor whose partial effect on y_i is of particular concern. In contrast, some or all elements of $x_{2,i}$ may be excluded from the model. Thus, $x_{2,i}$ may be a vector of control variables. The full model includes all regressors $x_i = (x_{1,i}', x_{2,i}')'$, while the reduced (smallest) model includes only $x_{1,i}$. There are up to 2^q possible combinations of regressors in the linear part.

Let $\mu = \mu(\beta)$ be our focus parameter, where $\mu(\cdot)$ is a known function. The focus parameter may merely be an element of β. Note that the focus parameter can depend on β_2, the coefficients of possibly excluded regressors. If elements of $x_{2,i}$ are excluded from the model, then the estimates of the corresponding coefficients are defined to be zero.

If $g(z)$ is known, then an FIC can be constructed in a way similar to that of Claeskens and Hjort (2003) (see Section 4.2). The full model, which is the true model, is defined as

$$y_i = x_{1,i}'\beta_1 + x_{2,i}'\delta_\beta/\sqrt{n} + g(z_i) + u_i, \quad E[u_i|x_i, z_i] = 0 \tag{2}$$

for some $\delta_\beta \neq 0$. Note that the model depends on the sample size n. Because (2) is the true model, a model is misspecified if some elements of $x_{2,i}$ are removed from the model. However, the misspecification is local in the sense that the model is within the $O(1/\sqrt{n})$ neighborhood of the true model. The reason for introducing the local misspecification framework is to consider the trade-off between the bias and the variance of the estimator for μ. If the true β_2 is a fixed vector, then the bias due to misspecification asymptotically dominates the variance of the estimator. That means that the full model is always the best model in terms of the asymptotic MSE. Under the local misspecification framework, the squared bias and the variance are both of the order $O(1/n)$.

When $g(z)$ is unknown, we need to determine the series length to approximate $g(z)$. Let $p^J(z) = (p^L(z)', p^K(z)')'$, where $p^L(z)$ and $p^K(z)$ are $L \times 1$ and $K \times 1$ vectors of basis functions, respectively. The vector $p^L(z)$ consists of basis functions that are definitely used to approximate $g(z)$, whereas some elements of $p^K(z)$ may not be used. In the case of a power series, for instance, $p^L(z)$ includes low-order power terms. The full model includes all basis functions, while the reduced model includes only $p^L(z_i)$. Then we consider the following full model:

$$y_i = x_{1,i}'\beta_1^* + x_{2,i}'\delta_\beta^*/\sqrt{n} + p^L(z_i)'\gamma_1^* + p^K(z_i)'\delta_\gamma^*/\sqrt{n} + u_i^*, \tag{3}$$

where u_i^* is orthogonal to x_i and $p^J(z_i)$. Thus, (3) represents the linear projection of y_i on $(x_i', p^J(z_i)')'$. That the coefficient of $p^K(z_i)$ is $O(1/\sqrt{n})$ describes the situation that $g(z)$ is sufficiently smooth and roughly approximated by $p^L(z_i)$.

Note that $\beta_1^* \neq \beta_1$ and $\delta_\beta^* \neq \delta_\beta$ in general. In contrast to (2), our full model (3) is misspecified in the sense that $E[u_i^*|x_i, z_i] \neq 0$, though it is valid as the linear projection. Thus, our setup is different from that of a standard FIC, which assumes that the full model is the true model. We consider $\beta^* = (\beta_1^{*\prime}, \delta_\beta^{*\prime}/\sqrt{n})'$ as the pseudo-true value that approximates the true parameter $\beta = (\beta_1', \delta_\beta'/\sqrt{n})'$. The closeness of the approximation depends on the series length in the full model and the smoothness of $g(z)$ and $E[x_i|z_i = z]$.

In the following section, we view $\mu^* = \mu(\beta^*)$ as an approximation of the true focus parameter $\mu = \mu(\beta)$ and we construct the FIC for μ^*. We simply assume that the difference between μ^* and μ is negligible. We do this by choosing sufficiently large J. This is partly due to the limitation of the series estimator. Unlike the kernel estimator, the bias of the series estimator is not obtained in the closed form. Only the order of the bias can be obtained, even if $g(z)$ is a known function.

It also makes sense to consider μ^* as the true focus parameter given that (2) is also an approximation that is introduced for mathematical convenience. In that case, our FIC is in spirit similar to the Takeuchi information criterion (Takeuchi, 1976).

3 Focused Information Criterion

Here, we derive the asymptotic distribution of the estimator and develop our FIC. We prepare some notations to describe candidate submodels. Let $x_{2j,i}$ and $p_j^K(z_i)$ be the j-th elements of $x_{2,i}$ and $p^K(z_i)$, respectively. Let S_q and S_K be subsets of $\{1,\ldots,q\}$ and $\{1,\ldots,K\}$. Also, let q_S and K_S denote the numbers of elements in S_q and S_K. We index a candidate model by a set $S=\{S_q,S_K\}$. The model S contains $x_{2j,i}$ and $p_l^K(z_i)$ as regressors if $j\in S_q$ and $l\in S_K$. The reduced and full models are the cases of $S=\{\emptyset,\emptyset\}$ and $S=\{\{1,\ldots,q\},\{1,\ldots,K\}\}$, respectively. Usually, the elements of $p^K(z_i)$ have a natural ordering. In that case, the number of models is at most $2^q\times(K+1)$.

Let $x_{2S,i}$ be the $q_S\times 1$ vector that contains $x_{2j,i}$, such that $j\in S_q$. Similarly, let $p_S^K(z_i)$ be the $K_S\times 1$ vector that contains $p_l^K(z_i)$, such that $l\in S_K$. Then the model S is

$$\begin{aligned} y_i &= x_{1,i}'\beta_{1,S}+x_{2S,i}'\beta_{2,S}+p^L(z_i)'\gamma_{1,S}+p_S^K(z_i)'\gamma_{2,S}+e_{Si} \\ &\equiv x_{S,i}'\beta_S+p_{S,i}'\gamma_S+e_{Si}. \end{aligned}$$

The estimator of β_S is

$$\hat{\beta}_S=\left(X_S'(I-Q_S)X_S\right)^{-1}X_S'(I-Q_S)Y,$$

where $Y=(y_1,\ldots,y_n)'$, $X_S=(x_{S,1},\ldots,x_{S,n})'$, $P_S=(p_{S,1},\ldots,p_{S,n})'$, and $Q_S=P_S(P_S'P_S)^{-1}P_S'$.

Let $x_{S^C,i}$ and $p_{S^C,i}$ be the vectors of excluded regressors from model S. Denote the residuals from the linear projection of $x_{S,i}$, $x_{S^C,i}$ and $p_{S^C,i}$ on $p_{S,i}$, by $\tilde{x}_{S,i}$, $\tilde{x}_{S^C,i}$ and $\tilde{p}_{S^C,i}$, respectively. Let $\Phi_S=E(\tilde{x}_{S,i}\tilde{x}_{S,i}')$, $\Psi_S=E(u_i^{*2}\tilde{x}_{S,i}\tilde{x}_{S,i}')$, and $B_S=(E(\tilde{x}_{S,i}\tilde{x}_{S^C,i}')\ E(\tilde{x}_{S,i}\tilde{p}_{S^C,i}'))$. We partition δ_β^* as $\delta_\beta^*=(\delta_{\beta,S}^{*'},\delta_{\beta,S^C}^{*'})'$, where $\delta_{\beta,S}^*$ and δ_{β,S^C}^* are the coefficient vectors of $x_{2S,i}$ and $x_{S^C,i}$ in the full model. Similarly, we write $\delta_\gamma^*=(\delta_{\gamma,S}^{*'},\delta_{\gamma,S^C}^{*'})'$ and define $\delta_S^*=(\delta_{\beta,S}^{*'},\delta_{\gamma,S}^{*'})'$ and $\delta_{S^C}^*=(\delta_{\beta,S^C}^{*'},\delta_{\gamma,S^C}^{*'})'$.

We assume that a standard regularity condition for the OLS estimator is satisfied. Then we have the following asymptotic results.

Lemma 1. *Suppose that $\{(y_i,x_i,z_i)\}_{i=1}^n$ are i.i.d. from (3). Then we have*

$$\sqrt{n}\begin{pmatrix}\hat{\beta}_{1,S}-\beta_1^*\\ \hat{\beta}_{2,S}\end{pmatrix}\xrightarrow{d}N\left(\begin{pmatrix}0_p\\ \delta_{\beta,S}^*\end{pmatrix}+\Phi_S^{-1}B_S\delta_{S^C}^*,\ \Phi_S^{-1}\Psi_S\Phi_S^{-1}\right)$$

as $n\to\infty$.

The proof is straightforward and so is omitted.

Note that J, the number of series terms in the full model, is fixed in our theory because we wish to find the optimal series length depending on the focus parameter. A standard asymptotic theory requires that $J \to \infty$ at a certain rate so that the bias due to nonparametric estimation is asymptotically negligible. This asymptotics does not suggest the determination of the number of series terms in practice. For instance, Zhang and Liang (2011) employed this asymptotics; therefore, their FIC is not applicable to the selection of the series terms. In contrast, we assume that our full model is sufficiently large so that the bias is rather small. Thus, we can capture the effect of using different series terms on the distribution of $\hat{\beta}_S$.

Let $\hat{\mu}_S = \mu(\hat{\beta}_S, 0_{q-q_S})$ be the estimator of μ^* based on model S. Note that the coefficients of $x_{S^C,i}$ are zero in model S. We define

$$A_S = \begin{pmatrix} E(\tilde{x}_{S,i}\tilde{x}'_{S^C,i}) & E(\tilde{x}_{S,i}\tilde{p}'_{S^C,i}) \\ -I_{q-q_S} & 0_{q-q_S \times K-K_S} \end{pmatrix}.$$

Lemma 1 and the delta method imply the following theorem.

Theorem 1. *Suppose that $\mu(\beta)$ is differentiable with respect to β. Then we have*

$$\sqrt{n}(\hat{\mu}_S - \mu^*) \xrightarrow{d} \left(\frac{\partial\mu}{\partial\beta}\right)' A_S \delta^*_{S^C} + \left(\frac{\partial\mu}{\partial\beta_S}\right)' N\left(0_{p+q_S}, \Phi_S^{-1}\Psi_S\Phi_S^{-1}\right),$$

where the derivatives $\partial\mu/\partial\beta$ and $\partial\mu/\partial\beta_S$ are evaluated at $(\beta_1^*, 0_q)$.

Again, the proof is omitted because it is straightforward.

By Theorem 1, the MSE of the limiting distribution of $\sqrt{n}(\hat{\mu}_S - \mu^*)$ is

$$\mathrm{mse}(S) = \left(\frac{\partial\mu}{\partial\beta}\right)' A_S \delta^*_{S^C}\delta^{*\prime}_{S^C} A'_S \left(\frac{\partial\mu}{\partial\beta}\right) + \left(\frac{\partial\mu}{\partial\beta_S}\right)' \Phi_S^{-1}\Psi_S\Phi_S^{-1}\left(\frac{\partial\mu}{\partial\beta_S}\right). \tag{4}$$

Owing to the local misspecification framework, the squared bias and the variance have the same order. In general, a large model implies a small first term and a large second term. The ranking of MSEs among different models is determined by the magnitude of the local parameters $\delta^* \equiv (\delta_\beta^{*\prime}, \delta_\gamma^{*\prime})'$. If $\delta^*/\sqrt{n}$ is sufficiently small, it would be better to set it at zero than to estimate it. Thus, the reduced model performs better than the full model, even though the reduced model is misspecified. Our FIC is obtained by replacing each component of (4) with its estimate.

We now discuss an estimation method for δ^*. Because of the nature of the local misspecification, we cannot consistently estimate δ^*. Let $(\hat{\beta}'_2, \hat{\gamma}'_2)'$ be the estimator of the coefficients of $w_i \equiv (x'_{2,i}, p^K(z_i)')'$ in the full model. Also, let $\tilde{w}_i$ be the residual of the linear projection of w_i on $(x'_{1,i}, p^L(z_i)')'$. We have the following result, which is similar to Lemma 1:

$$\hat{D} \equiv \sqrt{n}\begin{pmatrix} \hat{\beta}_2 \\ \hat{\gamma}_2 \end{pmatrix} \xrightarrow{d} D = N(\delta^*, \Gamma^{-1}\Omega\Gamma^{-1}),$$

where $\Gamma = E(\tilde{w}_i \tilde{w}_i')$ and $\Omega = E(u_i^{*2} \tilde{w}_i \tilde{w}_i')$. Because the mean of DD' is $\delta^* \delta^{*\prime} + \Gamma^{-1} \Omega \Gamma^{-1}$, we use $\widehat{\delta^* \delta^{*\prime}} = \hat{D}\hat{D}' - \hat{\Gamma}^{-1} \hat{\Omega} \hat{\Gamma}^{-1}$ to estimate $\delta^* \delta^{*\prime}$, where $\hat{\Gamma}$ and $\hat{\Omega}$ are sample analogs of Γ and Ω. The error term u_i^* can be replaced with the OLS residual from the full model. Note that u_i^* is heteroskedastic even if u_i is homoskedastic because u_i^* involves the approximation error. Thus, we need to use a heteroskedasticity-robust method to estimate the variance.

The matrices Φ_S, Ψ_S, and A_S are consistently estimated by their sample analogs. Therefore, the FIC of model S is defined as

$$\text{FIC}(S) = \left(\widehat{\frac{\partial \mu}{\partial \beta}} \right)' \hat{A}_S \widehat{\delta_{SC} \delta_{SC}'} \hat{A}_S' \left(\widehat{\frac{\partial \mu}{\partial \beta}} \right) + \left(\widehat{\frac{\partial \mu}{\partial \beta_S}} \right)' \hat{\Phi}_S^{-1} \hat{\Psi}_S \hat{\Phi}_S^{-1} \left(\widehat{\frac{\partial \mu}{\partial \beta_S}} \right). \tag{5}$$

The estimator for $\delta_{SC} \delta_{SC}'$ is the corresponding component of $\widehat{\delta^* \delta^{*\prime}}$. The derivatives are evaluated at $(\hat{\beta}_1, 0)$.

So far, we have considered the case in which the focus parameter depends only on β^*. It is also possible that the focus parameter depends on $\gamma^* = (\gamma_1^{*\prime}, \delta_\gamma^{*\prime}/\sqrt{n})'$. We briefly discuss an FIC for estimating the nonparametric part of the model. For some fixed value $\tilde{z} \in \mathbb{R}$, we define the focus parameter as $\mu(\gamma^*) = p^J(\tilde{z})'\gamma^* = p^L(\tilde{z})'\gamma_1^* + p^K(\tilde{z})'\delta_\gamma^*/\sqrt{n}$, which is estimated by $p_S(\tilde{z})'\hat{\gamma}_S$. Again, $\mu(\gamma^*)$ is an approximation of the true parameter of interest, $g(\tilde{z})$. The derivation of the FIC is almost the same as that of $\mu(\beta^*)$. The FIC is obtained by reversing the roles of x_i and $p^J(z_i)$ in (5).

4 Monte Carlo study

We investigate the performance of the FIC by means of a simple Monte Carlo study. We compare three model selection criteria: (i) FIC, (ii) AIC, and (iii) BIC. The true data generating process (DGP) is given by

$$y_i = x_{1,i}'\beta_1 + x_{2,i}'\delta_\beta/\sqrt{n} + g(z_i) + u_i,$$

where $\beta_1 = (\beta_{1,1}, \beta_{1,2})' = (1, 1)'$ and $\delta_\beta = (\delta_{\beta,1}, \dots, \delta_{\beta,4})'$. Each element of $x_{1,i}$ is independently generated from a standard normal distribution. The regressor z_i is uniformly distributed with support $[-2, 2]$. The regressor $x_{2j,i}$ is generated by $x_{2j,i} = 0.33 z_i + \epsilon_{ij}$ for $j = 1, \dots, 4$, where $\epsilon_{ij} \sim N(0, 2.25)$, so that some of the regressors in the linear part are correlated with z_i. The error term u_i is generated from a standard normal distribution and is independent of all regressors. We set $g(z) = \sin(1.5z)/(1.5 - \sin(1.5z))$. This function is sufficiently smooth, so it can be well approximated by the low-order polynomial function of z.

We estimate $g(z)$ by using power series. We use $p^L(z) = (1, z, \dots, z^4)'$ and $p^K(z) = (z^5, z^6, \dots, z^{11})'$. To simplify the description of candidate models, we consider nested models for the variable selection in the linear part. Thus, the number of candidate models is $5 \times 8 = 40$ in total.

We consider the following two DGPs: (1) $\delta_{\beta,j} = 1.5$ and (2) $\delta_{\beta,j} = 10(0.5^j)$ for $j = 1, \dots, 4$. The local parameter $\delta_{\beta,j}$ is constant for all j in the former case, whereas it declines as j increases in the latter case. The focus parameter is $\mu(\beta) = \sum_{j=1}^{2} \beta_{1,j} + \sum_{j=1}^{4} \delta_{\beta,j}/\sqrt{n}$.

To evaluate the performance of the selection criteria, we calculate the mean of the root MSE (RMSE) of each post-selection estimator based on the $5,000$ repetitions for three sample sizes: $n = 50$, 100, and 200. Note that the RMSEs are evaluated at the true focus parameter. We also report the RMSEs based on the reduced, middle, and full models, where the middle model is defined by $S = \{\{1, 2\}, \{1, 2, 3, 4\}\}$.

The result is summarized in Table 1. For DGP (1), we can see that the FIC outperforms other procedures for all sample sizes. The FIC successfully selects the model that has a lower MSE than the other candidate models. This result suggests that although our FIC is constructed for the approximate focus parameter, it works well even for the true focus parameter. For DGP (2), we also see that the FIC dominates other selection criteria. An important observation in DGP (2) is that the middle model is better than the full model. Thus, the full model is not always the best model. Also, the FIC-selected model performs better than the full model. Although the AIC and BIC tend to select relatively small models, the FIC can select proper models.

Next, we investigate the performance of the FIC for estimating the nonparametric component. The focus parameter is $g(1)$. Table 2 summarizes the result. For DGP (1), the performance of the BIC is slightly better than that of the other criteria for small sample sizes. The FIC achieves a lower MSE than the others for $n = 200$. A similar result is also obtained for DGP (2). This result confirms that our FIC can also be a useful tool for estimating the nonparametric part of the PLM.

Table 1: RMSE for $\mu(\beta)$

	DGP (1)			DGP (2)		
n	50	100	200	50	100	200
Full	0.626	0.387	0.259	0.620	0.392	0.260
Middle	0.633	0.422	0.292	0.469	0.311	0.216
Reduced	0.894	0.630	0.441	0.722	0.502	0.358
AIC	0.710	0.487	0.335	0.585	0.394	0.272
BIC	0.808	0.586	0.418	0.613	0.438	0.319
FIC	0.643	0.421	0.286	0.585	0.375	0.253

Table 2: RMSE for $g(1)$

	DGP (1)			DGP (2)		
n	50	100	200	50	100	200
Full	0.564	0.344	0.230	0.562	0.346	0.235
Middle	0.454	0.286	0.199	0.447	0.299	0.218
Reduced	0.344	0.286	0.288	0.370	0.319	0.312
AIC	0.523	0.365	0.261	0.506	0.361	0.268
BIC	0.440	0.327	0.287	0.451	0.361	0.318
FIC	0.514	0.331	0.240	0.518	0.340	0.245

5 Conclusion

In this paper, we propose an FIC for series estimation in the PLM. We approximate the true regression function with a linear projection of a finite-dimensional model, and derive the FIC that allows us to jointly select the variables in the linear part and the series length in the nonparametric part. This study is the first attempt to select the smoothing parameter by using the local misspecification framework in semiparametric models.

The simulation results are encouraging. Although our FIC is constructed for the approximate focus parameters, it works reasonably well even for the true focus parameters. Our FIC is a useful alternative tool for variable selection in the PLM.

Acknowledgements

The authors would like to thank Kengo Kato, Kosuke Oya, Eiji Kurozumi and conference participants at the 2013 Spring Meeting of the Japan Statistical Society at Gakushuin University for their helpful comments and suggestions. This research was partially supported by the Japan Society for the Promotion of Science under KAKENHI 23730215.

References

Akaike, H. (1973) "Information Theory and an Extension of the Maximum Likelihood Principle", in Petrov, B. and Csáki, F. eds., *Second International Symposium on Information Theory*, Akadémiai Kiadó, pp. 267–281.

Chen, H. (1988) "Convergence Rates for Parametric Components in a Partly Linear Model", *Annals of Statistics*, Vol. 16, pp. 136–146.

Claeskens, G., and N. L. Hjort (2003) "The Focused Information Criterion", *Journal of the American Statistical Association*, Vol. 98, pp. 900–916.

Ding, H., G. Claeskens and M. Jansen (2011) "Variable Selection in Partially Linear Wavelet Models", *Statistical Modelling*, Vol. 11, pp. 409–427.

Donald, S. G., and W. K. Newey (1994) "Series Estimation of Semilinear Models", *Journal of Multivariate Analysis*, Vol. 50, pp. 30–40.

Engel, R. F., C. W. J. Granger, J. Rice and A. Weiss (1986) "Semiparametric Estimates of the Relation between Weather and Electricity Demand", *Journal of the American Statical Association*, Vol. 81, pp. 310–320.

Härdle, W., H. Liang, and J. Gao (2013) *Patially Linear Models*: Springer.

Liang, H., W. Härdle and R. J. Carroll (1999) "Estimation in a Semiparametric Partially Linear Errors-in-Variable Model", *Annals of Statistics*, Vol. 27, pp. 1519–1535.

Liu, C.-A. (2013) "Distribution Theory of the Least Squares Averaging Estimator", National University of Singapore.

Mallows, C. L. (1973) "Some Comments on C_p", *Technometrics*, Vol. 15, pp. 661–675.

Robinson, P. M. (1988) "Root-N-Consistent Semiparametric Regression", *Econometrica*, Vol. 56, pp. 931–954.

Schwarz, G. (1978) "Estimating the Dimension of a Model", *Annals of Statistics*, Vol. 6, pp. 461–464.

Speckman, P. (1988) "Kernel Smoothing in Partial Linear Models", *Journal of the Royal Statistical Society, Ser. B*, Vol. 50, pp. 413–436.

Takeuchi, K. (1976) "Distribution of Information Statistics and Criteria for Adequacy of Models", *Suri Kagaku*, Vol. 153, pp. 12–18, in Japanese.

Zhang, H. H., G. Cheng and Y. Liu (2011) "Linear or Nonlinear? Automatic Structure Discovery for Partially Linear Models", *Journal of the American Statistical Association*, Vol. 106, pp. 1099–1112.

Zhang, X., and H. Liang (2011) "Focused Information Criterion and Model Averaging for Generalized Additive Partial Linear Models", *Annals of Statistics*, Vol. 39, pp. 174–200.