The focused information criterion for varying-coefficient partially linear measurement error models

HaiYing Wang[1], Xinjie Chen[2] and Nancy Flournoy[3]

October 30, 2014

[1] *Department of Mathematics and Statistics, University of New Hampshire, Durham, New Hampshire 03824, USA, Email: haiying.wang@unh.edu*
[2] *Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100190*, P. R. China, Email: chenxinjie@amss.ac.cn*
[3] *Department of Statistics, University of Missouri, Columbia, Missouri 65211, USA, Email: FlournoyN@Missouri.edu*

Abstract

Under general parametric models, Claeskens and Hjort (2003) proposed a focused information criterion (FIC) for model selection which emphasizes the accuracy of estimation for particular parameters of interest. This paper extends their framework to include a semi-parametric varying-coefficient partially linear model when covariates in both the parametric and the non-parametric parts are subject to measurement errors. We allow the covariance matrices of the measurement errors to be unknown and be estimated by replicated observations. Also, we derive the asymptotic properties of the frequentist model average (FMA) estimator for the model in consideration, which generalizes the results obtained by Wang *et al.* (2012). In addition to asymptotic properties, finite sample performance of the proposed methods are examined in a simulation study, and a data set obtained from Continuing Survey of Food Intakes by Individuals conducted by the U.S. Department of Agriculture's (CSFII) is considered.

Keywords: Focused information criterion, measurement errors, model averaging, model selection, semi-parametric models

1 Introduction

Model selection has been an important part of any statistical analysis, and many methods and criteria have been proposed in the literature for choosing the "best model". One type of criteria are based on the likelihood such as the Akaike information criterion (AIC) (Akaike, 1973) and the Bayesian information criterion (BIC) (Schwarz, 1978). The generalization of these criteria is restricted because the likelihood is often intractable when the candidate models are complicated, e.g. for semi-parametric models. Also, these criteria emphasize selecting the "best model" in the sense that the chosen model is closest to the true model. But such a model may not be the "best model" for other purposes such as parameter estimation and prediction. Another class of methods are motivated by minimizing the prediction errors, including Mallows' C_p (Mallows, 1973), Cross Validation (Stone, 1974), Generalized Cross Validation (Craven and Wahba, 1979), among others. These criteria concern the quality of prediction and may not be optimal for estimation.

Recently, Claeskens and Hjort (2003) proposed the focused information criterion (FIC) that emphasizes the quality of an estimator for a particular parameter of interest. For a pre-specified parameter, the FIC selects the model that minimizes the asymptotic risk of estimators obtained from sub-models. Since this criterion focuses on estimators of a parameter, it identifies the "best model" for estimating the parameter of interest. In addition, although FIC was proposed under full parametric models, the criterion itself does not require the likelihood, so it is possible to apply it to complicated models. For example, Hjort and Claeskens (2006) studied the FIC for the Cox hazard regression model; Zhang and Liang (2011) generalized the criterion for generalized additive partial linear models; Zhang *et al.* (2012) considered a tobit model with a non-zero threshold and considered non-quadratic loss functions; Peña *et al.* (2013) applied the FIC approach to estimate benchmark doses in a risk assessment study.

The purpose of this paper is to extend the definition of the FIC, broadening the scope of Claeskens and Hjort (2003)'s framework to include a semi-parametric varying-coefficient partially linear model when covariates in both the parametric and non-parametric components are measured with errors. Although Wang *et al.* (2012) extended Hjort and Claeskens (2003)'s investigation on the frequentist model average (FMA) in the varying-coefficient partially linear measurement error (VCPLE) model, they did not consider the FIC. Additionally, we consider a more general situation in which the covariates in the non-parametric component are also subject to measurement errors; in Wang *et al.* (2012) these covariates could be observed precisely. In addition, we assume now that the covariance matrices of the measurement errors are unknown.

Hjort and Claeskens (2003) showed that, by taking a weighted average on sub-model estimates, the accuracy of estimation can be improved further, and moreover, the uncertainty in model selection step can be incorporated for subsequent inference. They developed an asymptotic framework for the FMA estimation based on general parametric models. The idea of the FMA estimator has drawn a lot of attention in recent years and much progress has been made as seen in Hjort and Claeskens (2006), Claeskens and Carroll (2007), Claeskens and Hjort (2008), Schomaker *et al.* (2010), Zhang and Liang (2011), Zhang *et al.* (2012), Wang and Zou (2012), Wang *et al.* (2012), Schomaker (2012), and Schomaker and Heumann (2014) among others. This paper is also a study the FMA approach and its asymptotic properties are derived, which extends Wang *et al.* (2012)'s work to a more general case.

The remainder of the paper is organized as follows. In Section 2, we discuss the model setup and estimation method. In Section 3, we provide the main theoretical results. Results of simulation experiments along with an example based on real data are contained in Section 4. Section 5 concludes, and technical details are given in the appendix.

2 Model setup and estimation procedure

Assume that independent and identically distributed samples (Y_i, W_i, ζ_i, T_i), $i = 1, ..., n$ are taken form the following VCPLE model:

$$\begin{cases} Y &= X^\top \theta + Z^\top \alpha(T) + \varepsilon, \\ W &= X + U, \\ \zeta &= Z + V, \end{cases} \tag{1}$$

where Y is the response variable; (X, Z, T) are covariates; $\theta = (\beta^\top, \gamma^\top)^\top$ with β and γ being p and q dimensional coefficient vectors, respectively; $\alpha(\cdot) = \{\alpha_1(\cdot), ..., \alpha_r(\cdot)\}^\top$ is a vector of r unknown functions; ε is a random error with mean 0 and variance σ^2 which is independent of (X, Z, T). T is a one dimensional random variable for simplicity. Here, it is assumed that X and Z cannot be observed directly. Instead, their surrogates $W = X + U$ and $\zeta = Z + V$, respectively, are observed, with U and V being vectors of random errors with mean 0 and covariance matrices Σ_u and Σ_v, respectively. Furthermore, U and V are mutually independent and they are independent of (X, Z, T) and ε.

Wang *et al.* (2012) considered a similar model setup, but in their analysis Z is free of measurement errors and Σ_u is known. In this paper, we deal with the case that Σ_u and Σ_v are both unknown and replicated observations are available; that is, $W_{ij} = X_i + U_{ij}$ and $\zeta_{il} = Z_i + V_{il}$ are observed, $j = 1, ..., J$, $l = 1, ..., L$, $i = 1, ..., n$ (see Carroll *et al.*, 2006). To facilitate the discussion, write $\mathbf{Y} = (Y_1, ..., Y_n)^\top$, $\mathbf{X} = (X_1, ..., X_n)^\top$, $\mathbf{W} = (W_1, ..., W_n)^\top$, $\mathbf{U} = (U_1, ..., U_n)^\top$, $\mathbf{Z} = (Z_1, ..., Z_n)^\top$, $\mathbf{T} = (T_1, ..., T_n)^\top$, $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)^\top$ and $\mathbf{M} = \{Z_1^\top \alpha(T_1), ..., Z_n^\top \alpha(T_n)\}^\top$.

Now we discuss the estimation approach starting with the case when both X_i and Z_i are measured with errors with known Σ_u and Σ_v. In this case, the estimator given in Wang *et al.* (2012) is inconsistent because the measurement errors on Z_i are not taken into account. To ensure consistency, the following modified profile least-squares estimator is adopted:

$$\hat{\theta} = \arg\min_{\theta} \left[\sum_{i=1}^{n} \left\{ Y_i - \hat{Y}_i - (W_i - \hat{W}_i)^\top \theta \right\}^2 - n\theta^\top \Sigma_u \theta \right], \tag{2}$$

where $\hat{Y}_i = \psi_i \mathbf{Y}$, $\hat{W}_i = \{\psi_i \mathbf{W}\}^\top$, $\psi_i = (\boldsymbol{\zeta}_i^\top,\ 0) \left\{ (\mathcal{D}_{t_i}^{\zeta})^\top \Omega_{t_i} \mathcal{D}_{t_i}^{\zeta} - \phi_{t_i} \right\}^{-1} (\mathcal{D}_{t_i}^{\zeta})^\top \Omega_{t_i}$,

$$D_{t_i}^{\zeta} = \begin{pmatrix} \boldsymbol{\zeta}_1^\top & \frac{T_1 - t_i}{h} \boldsymbol{\zeta}_1^\top \\ \vdots & \vdots \\ \boldsymbol{\zeta}_n^\top & \frac{T_n - t_i}{h} \boldsymbol{\zeta}_n^\top \end{pmatrix}, \ \phi_{t_i} = \sum_{j=1}^{n} \begin{pmatrix} 1 & \frac{T_j - t_i}{h} \\ \frac{T_j - t_i}{h} & \frac{(T_j - t_i)^2}{h^2} \end{pmatrix} \otimes \Sigma_v K_h(T_j - t_i),$$

and $\otimes$ is the Kronecker product. The usage of $n\theta^\top \Sigma_u \theta$ in (2) is the so called "correction for attenuation" for measurement errors in the covariates of the linear part of the model (Liang *et al.*, 1999; You and Chen, 2006). The term ϕ_{t_i} is a correction suggested by You *et al.* (2006) for the varying-coefficient model under measurement errors. It has the purpose of correcting the bias introduced by measurement errors in the non-parametric part of the model. You *et al.* (2006) showed that the estimator of the unknown function under their model setup is inconsistent if this term is dropped. One can write equation (2) in matrix notation:

$$\hat{\theta} = \left(\widetilde{\mathbf{W}}^\top \widetilde{\mathbf{W}} - n\Sigma_u\right)^{-1} \widetilde{\mathbf{W}}^\top \widetilde{\mathbf{Y}}, \tag{3}$$

where $\widetilde{\mathbf{Y}} = (Y_1 - \hat{Y}_1, ..., Y_n - \hat{Y}_n)^\top$ and $\widetilde{\mathbf{W}} = (W_1 - \hat{W}_1, ..., W_n - \hat{W}_n)^\top$.

If Σ_u and Σ_v are unknown, they can be consistently and unbiasedly estimated by

$$\hat{\Sigma}_u = \frac{1}{n(J-1)} \sum_{i=1}^{n} \sum_{j=1}^{J} (W_{ij} - \bar{W}_i)^{\otimes 2} \text{ and } \hat{\Sigma}_v = \frac{1}{n(L-1)} \sum_{i=1}^{n} \sum_{l=1}^{L} (\zeta_{il} - \bar{\zeta}_i)^{\otimes 2},$$

respectively, where $\bar{W}_i = \sum_{j=1}^{J} W_{ij}/J$ and $\bar{\zeta}_i = \sum_{l=1}^{L} \zeta_{il}/L$. With the availability of replicated observations, the mean surrogates $\bar{W}_i$ and $\bar{\zeta}_i$ should be used since they have smaller measurement errors. Let $\bar{U}_i = \sum_{j=1}^{J} U_{ij}/J$, $\bar{V}_i = \sum_{l=1}^{L} V_{il}/L$ and $\bar{\mathbf{W}}$, $\bar{\mathbf{U}}$, $\bar{\boldsymbol{\zeta}}$ and $\bar{\mathbf{V}}$ be the matrices consisting of $\bar{W}_i$, $\bar{U}_i$, $\bar{\zeta}_i$ and $\bar{V}_i$, $i = 1,...,n$, respectively.

Using the means $\bar{W}_i$ and $\bar{\zeta}_i$, an estimator analogous to (3) is

$$\hat{\theta} = \left(\widetilde{\bar{\mathbf{W}}}^\top \widetilde{\bar{\mathbf{W}}} - nJ^{-1}\hat{\Sigma}_u\right)^{-1} \widetilde{\bar{\mathbf{W}}}^\top \widetilde{\bar{\mathbf{Y}}}, \tag{4}$$

where $\widetilde{\bar{\mathbf{Y}}} = (Y_1 - \hat{\bar{Y}}_1, ..., Y_n - \hat{\bar{Y}}_n)^\top$, $\widetilde{\bar{\mathbf{W}}} = (W_1 - \hat{\bar{W}}_1, ..., W_n - \hat{\bar{W}}_n)^\top$, $\hat{\bar{Y}}_i = \bar{\psi}_i \mathbf{Y}$, $\hat{\bar{W}}_i = \left\{\bar{\psi}_i \bar{\mathbf{W}}\right\}^\top$, $\bar{\psi}_i = (\bar{\boldsymbol{\zeta}}_i^\top,\ 0) \left\{(\mathcal{D}_{t_i}^{\bar{\zeta}})^\top \Omega_{t_i} \mathcal{D}_{t_i}^{\bar{\zeta}} - \bar{\phi}_{t_i}\right\}^{-1} (\mathcal{D}_{t_i}^{\bar{\zeta}})^\top \Omega_{t_i}$, $D_{t_i}^{\bar{\zeta}} = \begin{pmatrix} \bar{\boldsymbol{\zeta}}_1^\top & \frac{T_1 - t_i}{h} \bar{\boldsymbol{\zeta}}_1^\top \\ \vdots & \vdots \\ \bar{\boldsymbol{\zeta}}_n^\top & \frac{T_n - t_i}{h} \bar{\boldsymbol{\zeta}}_n^\top \end{pmatrix}$ and $\bar{\phi}_{t_i} = \frac{1}{L} \sum_{j=1}^{n} \begin{pmatrix} 1 & \frac{T_j - t_i}{h} \\ \frac{T_j - t_i}{h} & \frac{(T_j - t_i)^2}{h^2} \end{pmatrix} \otimes \hat{\Sigma}_v K_h(T_j - t_i)$.

3 FIC and model averaging

Following the local mis-specification framework used in Hjort and Claeskens (2003) and Wang *et al.* (2012), we let the true value of θ be $\theta_{\text{true}} = (\beta^\top, \gamma_{\text{true}}^\top)^\top = (\beta^\top, \delta^\top/\sqrt{n})^\top$, where the parameter γ represents the degree of a model's departure from the narrow model in which $\theta{=}\theta_0{=}(\beta^\top, 0^\top)^\top$.

3.1 Estimation of coefficients under the full model and sub-models

When $n \to \infty$, by an approach similar to the proof of Theorem 1 in Wang *et al.* (2013), we obtain $n^{-1}\widetilde{\bar{\mathbf{W}}}^\top \widetilde{\bar{\mathbf{W}}} \xrightarrow{p} J^{-1}\Sigma_u + B$, where

$$\begin{aligned} B =& \mathbf{E}\left(XX^\top\right) - \mathbf{E}\left[\mathbf{E}(XZ^\top|T)\left\{\mathbf{E}(ZZ^\top|T)\right\}^{-1}\mathbf{E}(XZ^\top|T)^\top\right] \\ &+ J^{-1}\mathbf{E}\left[\mathbf{E}(XZ^\top|T)\left\{\mathbf{E}(ZZ^\top|T)\right\}^{-1}\Sigma_v\left\{\mathbf{E}(ZZ^\top|T)\right\}^{-1}\mathbf{E}(XZ^\top|T)^\top\right], \end{aligned}$$

and $\xrightarrow{p}$ denotes convergence in probability. Accordingly, a consistent estimator of B is $\hat{B} = n^{-1}\widetilde{\bar{\mathbf{W}}}^\top\widetilde{\bar{\mathbf{W}}} - J^{-1}\hat{\Sigma}_u \equiv n^{-1}B_n$. If Σ_u is known and no replicates are available, then $\widetilde{\bar{\mathbf{W}}}$ and $J^{-1}\hat{\Sigma}_u$ in $\hat{B}$ need to be replaced by $\widetilde{\mathbf{W}}$ and Σ_u, respectively.

Following an approach similar to that of Wang *et al.* (2012), we obtain the relationship between the estimator under the full, $(\hat{\beta}_{\text{full}}^\top, \hat{\gamma}_{\text{full}}^\top)^\top$, and the estimator under the sub-model S, $(\hat{\beta}_{\text{S}}^\top, \hat{\gamma}_{\text{S}}^\top)^\top$:

$$\begin{pmatrix}\hat{\beta}_{\text{S}}\\ \hat{\gamma}_{\text{S}}\end{pmatrix} = \begin{pmatrix} I_p & C_{ns}\\ 0_{|S|\times p} & (\Pi_{\text{S}}^\top A_n \Pi_{\text{S}})^{-1}\Pi_{\text{S}}^\top A_n\end{pmatrix}\begin{pmatrix}\hat{\beta}_{\text{full}}\\ \hat{\gamma}_{\text{full}}\end{pmatrix} \equiv G_{ns}\begin{pmatrix}\hat{\beta}_{\text{full}}\\ \hat{\gamma}_{\text{full}}\end{pmatrix}, \tag{5}$$

where $A_n = B_{n22} - B_{n21}B_{n11}^{-1}B_{n12}$; $\Pi_{\text{S}}^\top$ is an $|S| \times q$ selection matrix; $|S|$ is the number of elements in S; $C_{ns} = B_{n11}^{-1}B_{n12}\left(I_q - A_n^{-1/2}H_{ns}A_n^{1/2}\right)$; $H_{ns} = A_n^{1/2}\Pi_{\text{S}}\left(\Pi_{\text{S}}^\top A_n \Pi_{\text{S}}\right)^{-1}\Pi_{\text{S}}^\top A_n^{1/2}$.

The following theorem illustrates the asymptotic properties of sub-model estimators.

Theorem 1. *If conditions 1-5 in the Appendix hold, and U_i, V_i, ε_i and (X_i, Z_i, T_i) are independent, then, as $n \to \infty$,*

$$\sqrt{n}\begin{pmatrix}\hat{\beta}_{\text{S}} - \beta\\ \hat{\gamma}_{\text{S}}\end{pmatrix} \xrightarrow{d} N\left\{\begin{pmatrix} C_{\text{S}}\delta\\ (\Pi_{\text{S}}^\top A \Pi_{\text{S}})^{-1}\Pi_{\text{S}}^\top A\delta\end{pmatrix},\ G_{\text{S}}PG_{\text{S}}^\top\right\};$$

specifically for estimator under the full model,

$$\sqrt{n}\begin{pmatrix}\hat{\beta}_{\text{full}} - \beta\\ \hat{\gamma}_{\text{full}}\end{pmatrix} \equiv \begin{pmatrix}M_n\\ \hat{\delta}\end{pmatrix} \xrightarrow{d} \begin{pmatrix}M\\ D\end{pmatrix} \sim N\left\{\begin{pmatrix}0\\ \delta\end{pmatrix},\ P\right\},$$

where $\xrightarrow{d}$ denotes convergence in distribution, $A = B_{22} - B_{21}B_{11}^{-1}B_{12}$, C_{S} and G_{S} are respectively the limits of A_n/n, C_{ns} and G_{ns}, $P = B^{-1}FB^{-1}$,

$$F = \mathbf{E}\left(\left[\bar{W}_i - \mathbf{E}\left(X_iZ_i^\top|T_i\right)\left\{\mathbf{E}\left(Z_iZ_i^\top|T_i\right)\right\}^{-1}\bar{\zeta}_i\right]\left(\varepsilon_i - \bar{U}_i^\top\theta_0\right) + \frac{\sum_{j=1}^J(U_{ij} - \bar{U}_i)^{\otimes 2}\theta_0}{J(J-1)}\right)^{\otimes 2},$$

and $R^{\otimes 2} = RR^\top$ for any matrix R.

From the proof in the appendix, a consistent estimator of F is

$$\hat{F} = \frac{1}{n}\sum_{i=1}^n\left[\left(\bar{W}_i - \hat{\bar{W}}_i\right)\left\{\left(Y_i - \hat{\bar{Y}}_i\right) - \left(\bar{W}_i - \hat{\bar{W}}_i\right)^\top\hat{\theta}\right\} + \frac{\sum_{j=1}^J(W_{ij} - \bar{W}_i)^{\otimes 2}\hat{\theta}}{J(J-1)}\right]^{\otimes 2},$$

where $\hat{\theta} = (\hat{\beta}_{\text{full}}^\top, \hat{\gamma}_{\text{full}}^\top)^\top$. Hence the asymptotic variance $G_{\text{S}}PG_{\text{S}}^\top$ can be consistently estimated by $G_{ns}\hat{P}G_{ns}^\top$, where $\hat{P} = \hat{B}^{-1}\hat{F}\hat{B}^{-1}$.

3.2 Main results

In this subsection, we consider the estimation of the parameter $\mu_{\text{true}} = \mu(\beta, \gamma_{\text{true}})$. The focused parameter does not depend on variance components as we assume the same variance structure for all models. Let the estimator based on sub-model S be $\hat{\mu}_{\rm S} = \mu(\hat{\beta}_{\rm S}, \hat{\gamma}_{\rm S})$; the estimate is inserted if the corresponding coefficient is included in the model and 0 is used otherwise. The following theorem is obtained.

Theorem 2. *Assume that μ is differentiable at $\theta_0 = (\beta^\top, 0^\top)^\top$. If conditions 1-5 in the Appendix are satisfied, and U_i, V_i, ε_i and (X_i, Z_i, T_i) are independent, then*

$$\sqrt{n}(\hat{\mu}_{\rm S} - \mu_{\rm true}) \xrightarrow{d} \Lambda_{\rm S} = \mu_\beta^\top \left\{M + B_{11}^{-1}B_{12}(D-\delta)\right\} + \omega^\top \left\{\delta - A^{-1/2}H_{\rm S}A^{1/2}D\right\},$$

where $\omega = B_{21}B_{11}^{-1}\mu_\beta - \mu_\gamma$, $\mu_\beta = \partial\mu(\beta,0)/\partial\beta$, $\mu_\gamma = \partial\mu(\beta,0)/\partial\gamma$, and $H_{\rm S}$ has the same form as H_{ns} with A_n being replaced by A.

From Theorem 2, the asymptotic mean square error of $\hat{\mu}_{\rm S}$ is

$$\mathbf{E}(\Lambda_{\rm S}^2) = \omega^\top(I_q - A^{-1/2}H_{\rm S}A^{1/2})\delta\delta^\top(I_q - A^{-1/2}H_{\rm S}A^{1/2})^\top\omega^\top + (\mu_\beta^\top, \mu_\gamma^\top\Pi_{\rm S})G_{\rm S}PG_{\rm S}^\top(\mu_\beta^\top, \mu_\gamma^\top\Pi_{\rm S})^\top.$$

$\mathbf{E}(\Lambda_{\rm S}^2)$ can be unbiasedly estimated by

$$\begin{aligned}\widehat{r(S)} =& \omega^\top(I_q - A^{-1/2}H_{\rm S}A^{1/2})\{DD^\top - (0, I_q)P(0, I_q)^\top\}(I_q - A^{-1/2}H_{\rm S}A^{1/2})^\top\omega \\ &+ (\mu_\beta^\top, \mu_\gamma^\top\Pi_{\rm S})G_{\rm S}PG_{\rm S}^\top(\mu_\beta^\top, \mu_\gamma^\top\Pi_{\rm S})^\top \\ =& \left\{\omega^\top(I_q - A^{-1/2}H_{\rm S}A^{1/2})D\right\}^2 + (\mu_\beta^\top, \mu_\gamma^\top + 2\omega^\top)P(\mu_\beta^\top, \mu_\gamma^\top)^\top \\ &- 2(0, \omega^\top A^{-1/2}H_{\rm S}A^{1/2})P(\mu_\beta^\top, \mu_\gamma^\top)^\top.\end{aligned}$$

Following the idea of Claeskens and Hjort (2003), we drop the constant term $(\mu_\beta^\top, \mu_\gamma^\top + 2\omega^\top)P(\mu_\beta^\top, \mu_\gamma^\top)^\top$ and define the theoretical FIC value for this model to be

$$\text{FIC}_{\rm S} = \left\{\omega^\top(I_q - A^{-1/2}H_{\rm S}A^{1/2})D\right\}^2 - 2(0, \omega^\top A^{-1/2}H_{\rm S}A^{1/2})P(\mu_\beta^\top, \mu_\gamma^\top)^\top. \tag{6}$$

Note that if there is no measurement error, then the second term (including the negative sign) on the right hand side of (6) reduces to $2\omega^\top\Pi_{\rm S}(\Pi_{\rm S}^\top A\Pi_{\rm S})^{-1}\Pi_{\rm S}^\top\omega$, so the FIC for the varying-coefficient partially linear model has the same expression as that defined in Claeskens and Hjort (2003) for parametric models. This FIC is for the limit experiment. For practical analysis, we need to plug in estimates for unknown parameters and thereby obtain a definition for the real FIC.

Definition 1. *The real FIC for the VCPLE model is defined as*

$$FIC_{ns} = \left\{\hat{\omega}^\top(I_q - A_n^{-1/2}H_{ns}A_n^{1/2})\hat{\delta}\right\}^2 - 2(0, \hat{\omega}^\top A_n^{-1/2}H_{ns}A_n^{1/2})\hat{P}_n(\hat{\mu}_\beta^\top, \hat{\mu}_\gamma^\top)^\top,$$

where $\hat{\omega}$, $\hat{\mu}_\beta$ and $\hat{\mu}_\gamma$ are consistent estimates of ω, μ_β and μ_γ, respectively. For model selection, the sub-model with the smallest value of FIC_{ns} is selected.

To gain further efficiency and to take into account the variation from the model selection stage, we also consider the frequentist model averaging (FMA) estimator as in Hjort and Claeskens (2003) and Wang *et al.* (2012). The FMA estimator considered here has the following form:

$$\hat{\mu}_{\rm avg} = \sum_{\rm S} c(S|\hat{\delta})\hat{\mu}_{\rm S}, \tag{7}$$

where $c(S|\hat{\delta})$'s are weight functions that sum to one. Theorem 3 depicts the asymptotic properties of $\hat{\mu}_{\text{avg}}$.

Theorem 3. *Assume that μ is differentiable at θ_0, and the weight functions $c(S|d)$'s are continuous almost everywhere. If conditions 1-5 in the Appendix hold, and U_i, V_i, ε_i and (X_i, Z_i, T_i) are independent, then we have*

$$\begin{aligned}\sqrt{n}(\hat{\mu}_{\text{avg}} - \mu_{\text{true}}) \xrightarrow{d} \Lambda = \mu_\beta^\top \left\{M + B_{11}^{-1}B_{12}(D-\delta)\right\} + \omega^\top \left\{\delta - Q(D)D\right\}, \\ \mathbf{E}\Lambda = \omega^\top \left[\delta - \mathbf{E}\{Q(D)D\}\right], \ \textit{and} \\ \mathbf{Var}(\Lambda) = \mu_\beta^\top (I, B_{11}^{-1}B_{12})P(I, B_{11}^{-1}B_{12})^\top \mu_\beta + \omega^\top \mathbf{Var}\left\{Q(D)D\right\}\omega \\ - 2\mu_\beta^\top (I, B_{11}^{-1}B_{12})\mathbf{Cov}\left\{(M^\top, D^\top)^\top, Q(D)D\right\}\omega,\end{aligned}$$

where $Q(D) = A^{-1/2}\left\{\sum_{\text{S}} c(S|D)H_{\text{S}}\right\}A^{1/2}$.

The limiting distribution in Theorem 3 is not normal since the weights are functions of the random variable D. Confidence intervals of unknown parameters can be constructed analogously to equation (12) in Wang *et al.* (2012).

4 Numerical studies

4.1 Simulation studies

In this section, we conduct simulation experiments to evaluate the finite sample performance of the FIC and the FMA approach. We simulate samples from the VCPLE model (1) for the case when $\alpha(T) = \{\sin(6\pi T), \sin(2\pi T)\}^\top$. The covariates, random error and the measurement errors are generated as follows. Covariates X and Z are generated from $N(0, I_6)$ and $N(0, I_2)$, respectively; T is generated from a uniform distribution on $(0, 1)$; ε follows a standard normal distribution; measurement errors U and V are generated from normal distributions $N(0, \sigma_u^2 I_6)$ and $N(0, \sigma_v^2 I_2)$, respectively. To estimate the covariance matrices of U and V, each four replicates of W and ζ are generated (i.e. $J = L = 4$). For the true value of each parameter, we let $\sigma_u = \sigma_v = 0.1, 0.5$ and $\theta = \{\beta^\top, \gamma^\top\}^\top = \{(1.5, 2), \delta^\top/\sqrt{n}\}^\top$. We consider three cases of δ: $\delta^{(1)} = (0, 0, 0, 0)^\top$, $\delta^{(2)} = (0.5, 0, 0.5, 0)^\top$ and $\delta^{(3)} = (0.5, 0.5, 0.5, 0.5)^\top$. We focus our interest on two estimands, $\mu_1 = \beta_1$ and $\mu_2 = \beta_1 + \beta_2 + \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4$.

Following Buckland *et al.* (1997)'s suggestion, we set the weights based on smoothed AIC (S-AIC), smoothed BIC (S-BIC) and smoothed FIC (S-FIC) values as $\frac{\exp(-\text{AIC}_{ns}/2)}{\sum_S \exp(-\text{AIC}_{ns}/2)}$, $\frac{\exp(\text{BIC}_{ns}/2)}{\sum_S \exp(-\text{BIC}_{ns}/2)}$, and $\frac{\exp(-\text{FIC}_{ns}/2)}{\sum_S \exp(-\text{FIC}_{ns}/2)}$ respectively. We compare these FMA estimators with the estimators obtained from the full model along with AIC-, BIC- and FIC-based model selection, and we use empirical mean square error (MSE) to evaluate the performance of estimators. Empirical MSEs are calculated based on 1000 independent samples of sizes $n = 100$ and $n = 200$.

Table 1 presents the MSEs of each estimator relative to that of the full model. A relative empirical MSE value smaller than 1 indicates that the given method is superior to the full model estimator, and vice versa. From the results we have the following discoveries. First, FIC based methods are superior to the corresponding AIC or BIC based method in term of MSE. This agrees with the fact that FIC aims to select a model with minimum asymptotic MSE. For FIC model averaging and FIC model selection, no method dominates the other, and their relative performance depends on the parameter of interest as well as the true value of δ. The superiority of FIC based methods are more significant for μ_2 than they are for μ_1. Second, the S-AIC and S-BIC model averaging estimators always produce smaller MSEs than AIC and BIC model selection estimators, respectively. This pattern is obvious particularly if the parameter of interest is μ_2. Third, model averaging and model selection approach are generally better strategies than the full model approach. The only exception in our simulations is when the sample size $n = 100$ and our focus is on μ_1. In this scenario, AIC model selection estimator performs similarly to the full model estimator.

4.2 A real data example

As an illustration, we consider an application of the proposed methods to a subset of data from the Continuing Survey of Food Intakes by Individuals (CSFII) conducted by the U.S. Department of Agriculture. This is part of the Nationwide Food Consumption Survey, published by the U.S. Department of Agriculture Human Nutrition Information Service, Hyattsville, Maryland (CSFII Reports No. 85-4 and No. 86-3). The same data set was used in Wang *et al.* (2012), but they only investigated the FMA approach. Here we apply the FIC to select models in addition to using FMA estimation. Additionally, we include the race information in our analysis which was not considered in Wang *et al.* (2012). This data set contains dietary intake and related information of $n = 1827$ individuals between the age of 25 and 50. There were 36 individuals who did not provide their race, so we removed observations from these individuals and only used the remaining 1791 observations in our analysis. Using the available data, we specify the following model for calories intake, y:

$$y = \sum_{i=1}^{11} \beta_i x_i + f_0(t) + z f_1(t) + \varepsilon,$$

where x_1, x_2 and x_3 are intake levels of fat, protein, carbohydrates, respectively; x_4 is an indicator variable for alcohol consumers with value 1 for alcohol consumers and 0 otherwise; x_5 is the body mass index; x_6 and x_7 are intake levels of Vitamin C and Vitamin A respectively; x_8, x_9, x_{10} and x_{11} are indicator variables representing race categories White, Black, Asian and Aleut, respectively; z is income; t is age. In addition, x_6 and x_7 are measured with errors, and they are replaced by the mean values of the observed surrogates.

Following the idea of distinguishing between mandatory and optional explanatory variables in Magnus and Durbin (1999) and Danilov and Magnus (2004), we treat x_1, x_2 and x_3 as mandatory in the parametric component of the model because fat, protein and carbohydrates are the key determinants of calories and we are primarily interested in the effects that these variables have on calorie intake. We are less interested in the effects of other variables on y, so we treat them as optional.

We focus on five parameters of interest: $\mu_1 = \beta_1$, $\mu_2 = \beta_2$, $\mu_3 = \beta_3$, $\mu_4 = \sum_{i=1}^{11} \beta_i$ and $\mu_5 = \beta_1/\beta_2$. μ_1, μ_2 and μ_3 are the marginal effects that each of the mandatory explanatory variables have on calorie intake; μ_4 is an example of a linear combination of the marginal effects; μ_5 is of interest because it measures the effect of fat relative to that of protein. Seven estimation methods are considered: FMA by S-AIC, S-BIC and S-FIC, model selection by AIC, BIC and FIC, and full model estimation.

Tables 2 presents the estimation results, in which the numbers in parenthesis are 95% confidence intervals. For point estimates of the marginal effects, all methods produced estimates of μ_3 larger than that of μ_1 and μ_2, indicating that carbohydrates is the main calorie intake, followed by fat and protein. FIC based methods yield estimates of μ_1 that are larger and estimates of μ_2 that are smaller than the corresponding estimates obtained from other methods, which most accentuates the common belief that calorie intake is associated with fat consumption more than with other consumptions. As for interval estimation comparisons, note that for μ_1, μ_2, μ_3 and μ_4, model averaging and the full model estimation produce the same interval estimates as these estimands are all linear functions of regression coefficients. Also, model selection generally results in narrower confidence intervals than do full model estimation or model averaging. This should not be interpreted as that model selection produces more precise interval estimation because model selection neglects the uncertainty in the stage of model selection when constructing confidence intervals. For the purpose of comparison, results ignoring the measurement errors are ignored are also given in Table 2. In general, compared with the results when measurement errors are taken into account, the estimates of μ_1, μ_4 and μ_5 are larger, while the estimates of μ_2 and μ_3 are smaller. A particular interesting observation from the comparisons is that the difference is the least for the results using the FIC. The results are identical for μ_1 and μ_5. This does not mean that the FIC can remove the effects of measurement errors. The reason for this will be explained by the results of model selection below.

Table 3 gives the results for model selection. In addition to the mandatory variables, the AIC included three optional variables in the selected model: alcohol usage, Vitamin A intake level and one indicator of race categories. The BIC also included the alcohol usage and vitamin A intake level in the chosen model but exclude the indicator of race categories. The AIC and the BIC always chose the same model regardless of the parameter of interest. On the other hand, for different parameters of interest, the FIC selected different models. If the measurement errors are ignored, the AIC added vitamin C intake level to the model, while the BIC removed vitamin A level from the model. The FIC, on the other hand, selected the same models for estimating μ_1, μ_4 and μ_5. Furthermore, the models for estimating μ_1 and μ_5 do not include x_6 or x_7, the variables measured with errors, and that is the reason why the estimation results are identical for these parameters with or without measurement errors taken into account. We see that although the FIC cannot remove the effect of measurement errors, it may avoid the effect by selecting a suitable model that is not affected by the measurement errors for some parameters of interest.

5 Concluding remarks

In this paper, we derive the FIC for the varying-coefficient partially linear model when covariates are measured with errors, which has generalized Claeskens and Hjort (2003)'s frame work to include a larger class of models. Moreover, we have extended Wang *et al.* (2012)'s results to address a situation when covariates in both the parametric and the non-parametric parts of the varying-coefficient partially linear model are measured with errors and covariance matrices of measurement errors are unknown. We notice that the limiting variables Λ_{S} and Λ have the same expressions as those in Wang *et al.* (2012), but the underlying distributions are different. If Z_i is free of measurement errors, then our results reduce to those in Wang *et al.* (2012).

Acknowledgment

We are very grateful to the editor and two referees for their constructive comments and suggestions.

Appendix

The following conditions are required for the proof.

1. The random variable T has bounded support $\mathbf{\Omega}$, and its density f is Lipschitz continuous and bounded away from 0 on its support.
2. For each $T \in \mathbf{\Omega}$, the $r \times r$ matrix $\mathbf{E}(ZZ^\top|T)$ is non-singular, and each element of $\mathbf{E}(ZZ^\top|T)$, $\mathbf{E}(XX^\top|T)$ or $\mathbf{E}(ZX^\top|T)$ is Lipschitz continuous.
3. There exists some $\epsilon > 2$ such that $\mathbf{E}\|X\|^{2\epsilon} < \infty$, $\mathbf{E}\|Z\|^{2\epsilon} < \infty$, $\mathbf{E}\|U\|^{2\epsilon} < \infty$, $\mathbf{E}\|V\|^{2\epsilon} < \infty$ and $\mathbf{E}\|\varepsilon\|^{2\epsilon} < \infty$, and $\rho < 2 - \epsilon^{-1}$ such that $nh^{2\rho-1} \to \infty$ and $nh^8 \to 0$.
4. $\alpha_j(T), j = 1, ..., r$, is twice continuously differentiable in $T \in \mathbf{\Omega}$.
5. $K(\cdot)$ is a symmetric density with compact support.

Proof of Theorem 1. Let $\hat{U}_i = (\bar{\psi}_i \mathbf{U})^\top$, $\hat{\varepsilon}_i = \bar{\psi}_i \boldsymbol{\varepsilon}$ and $\bigtriangledown = \widetilde{\bar{\mathbf{W}}}^\top \widetilde{\bar{\mathbf{W}}} - nJ^{-1}\hat{\Sigma}_u$. Then from equation (4), we have

$$
\begin{aligned}
\hat{\theta} - \theta_{\text{true}} &= \bigtriangledown^{-1} \sum_{i=1}^{n} (\bar{W}_i - \hat{\bar{W}}_i)(Y_i - \hat{\bar{Y}}_i) - \bigtriangledown^{-1} \bigtriangledown \theta_{\text{true}} \\
&= \bigtriangledown^{-1} nJ^{-1} \hat{\Sigma}_u \theta_{\text{true}} + \bigtriangledown^{-1} \sum_{i=1}^{n} (\bar{W}_i - \hat{\bar{W}}_i) \left\{ Y_i - \hat{\bar{Y}}_i - (\bar{W}_i - \hat{\bar{W}}_i)^\top \theta_{\text{true}} \right\}.
\end{aligned}
$$

From the expressions of $\hat{\bar{Y}}_i$, $\hat{\bar{W}}_i$, $\hat{\bar{U}}_i$ and $\hat{\bar{\varepsilon}}_i$, we obtain

$$Y_i - \hat{\bar{Y}}_i - (\bar{W}_i - \hat{\bar{W}}_i)^\top \theta_{\text{true}} = Z_i^\top \alpha(T_i) + \varepsilon_i - \bar{U}_i^\top \theta_{\text{true}} - \hat{\varepsilon}_i + \hat{\bar{U}}_i^\top \theta_{\text{true}} - \bar{\psi}_i \mathbf{M}.$$

Hence,

$$\begin{aligned}
&\sum_{i=1}^n (\bar{W}_i - \hat{\bar{W}}_i) \left\{ Y_i - \hat{\bar{Y}}_i - (\bar{W}_i - \hat{\bar{W}}_i)\theta_{\text{true}} \right\} \\
=&\sum_{i=1}^n (\bar{W}_i - \hat{\bar{W}}_i)(\varepsilon_i - \bar{U}_i^\top \theta_{\text{true}}) + \sum_{i=1}^n (\bar{W}_i - \hat{\bar{W}}_i)(\hat{\bar{U}}_i^\top \theta_{\text{true}} - \hat{\varepsilon}_i) + \sum_{i=1}^n (\bar{W}_i - \hat{\bar{W}}_i)\{Z_i^\top \alpha(T_i) - \bar{\psi}_i \mathbf{M}\} \\
=&\sum_{i=1}^n \left[\bar{W}_i - \mathbf{E}(\bar{W}_i Z_i^\top | T_i) \{\mathbf{E}(Z_i Z_i^\top | T_i)\}^{-1} \bar{\zeta}_i \right] (\varepsilon_i - \bar{U}_i^\top \theta_{\text{true}}) \\
&+ \sum_{i=1}^n \left[\mathbf{E}(\bar{W}_i Z_i^\top | T_i) \{\mathbf{E}(Z_i Z_i^\top | T_i)\}^{-1} \bar{\zeta}_i - \hat{\bar{W}}_i \right] (\varepsilon_i - \bar{U}_i^\top \theta_{\text{true}}) \\
&+ \sum_{i=1}^n (\bar{W}_i - \hat{\bar{W}}_i)(\hat{\bar{U}}_i^\top \theta_{\text{true}} - \hat{\varepsilon}_i) + \sum_{i=1}^n (\bar{W}_i - \hat{\bar{W}}_i)\{Z_i^\top \alpha(T_i) - \bar{\psi}_i \mathbf{M}\} \\
\equiv& J_1 + J_2 + J_3 + J_4.
\end{aligned}$$

Applying the method used in Fan and Huang (2005) provides that, uniformly in T, $\hat{\bar{W}}_i^\top = (\bar{\boldsymbol{\zeta}}_i^\top,\ 0) \left\{ (\mathcal{D}_{t_i}^{\bar{\zeta}})^\top \Omega_{t_i} \mathcal{D}_{t_i}^{\bar{\zeta}} - \bar{\phi}_{t_i} \right\}^{-1} (\mathcal{D}_{t_i}^{\bar{\zeta}})^\top \Omega_{t_i} \mathbf{W} = \bar{\zeta}_i^\top \{\mathbf{E}(Z_i Z_i^\top | T_i)\}^{-1} \mathbf{E}(Z_i X_i^\top | T_i)\{1 + O_P(c_n)\}$, where $c_n = \{\log(1/h)/(nh)\}^{1/2} + h^2$. In addition, since $\{\mathbf{E}(\bar{W}_i Z_i^\top | T_i)\}^\top = \mathbf{E}(Z_i \bar{W}_i^\top | T_i)$ and $\theta_{\text{true}} = \theta_0 + (0^\top, \delta^\top)^\top / \sqrt{n}$, we have

$$J_2 = \sum_{i=1}^n [\mathbf{E}(\bar{W}_i Z_i^\top | T_i)\{\mathbf{E}(Z_i Z_i^\top | T_i)\}^{-1} \bar{\zeta}_i](\varepsilon_i - \bar{U}_i^\top \theta_0) O_P(c_n).$$

The application of the Central Limit Theorem yields $\sum_{i=1}^n [\mathbf{E}(\bar{W}_i Z_i^\top | T_i)\{\mathbf{E}(Z_i Z_i^\top | T_i)\}^{-1} Z_i](\varepsilon_i - \bar{U}_i^\top \theta_0) = O_P(\sqrt{n})$. Therefore, $J_2 = O_P(\sqrt{n} c_n) = o_P(\sqrt{n})$. Similarly $J_3 = o_P(\sqrt{n})$, and $J_4 = o_P(\sqrt{n})$. Using Slutsky's Theorem and recognizing that $\bigtriangledown / n = B_n \xrightarrow{p} B$ as $n \to \infty$, we obtain

$$\begin{aligned}
\sqrt{n}(\hat{\theta} - \theta_{\text{true}}) =& \left(\frac{\bigtriangledown}{n} \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \left(\bar{W}_i - E(\bar{W}_i Z_i^\top | T_i) \{E(Z_i Z_i^\top | T_i)\}^{-1} \bar{\zeta}_i \right) \right. \\
& \left. \times \left(\varepsilon_i - \bar{U}_i^\top \theta_{\text{true}} \right) + \frac{\sum_{j=1}^J (W_{ij} - \bar{W}_i)^{\otimes 2} \theta_{\text{true}}}{J(J-1)} \right\} + o_P(1) \\
=& \left(\frac{\bigtriangledown}{n} \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \left(\bar{W}_i - E(\bar{W}_i Z_i^\top | T_i) \{E(Z_i Z_i^\top | T_i)\}^{-1} \bar{\zeta}_i \right) \right. \\
& \left. \times (\varepsilon_i - \bar{U}_i^\top \theta_0) + \frac{\sum_{j=1}^J (W_{ij} - \bar{W}_i)^{\otimes 2} \theta_0}{J(J-1)} \right\} + o_P(1) \\
\xrightarrow{d}& N(0,\ B^{-1} F B^{-1}).
\end{aligned}$$

The above results together with equation (5) and the Continuous Mapping Theorem finish the proof. □

Proofs of Theorem 2 and Theorem 3. With the result in Theorem 1, they can be proved using approaches similar to those used in the proof of Theorem 1 and Theorem 2 in Wang *et al.* (2012), respectively. We skip the details here to save space. □

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory* (B. N. Petrov and F. Cźaki, eds), pp. 267–281. Akademiai Kaidó, Budapest.

Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics* **53**, 603–618.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective.* Chapman and Hall, New York, 2nd edn.

Claeskens, G. and Carroll, R. J. (2007). An asymptotic theory for model selection inference in general semiparametric problems. *Biometrika* **94**, 249–265.

Claeskens, G. and Hjort, N. L. (2003). The focused information criterion. *Journal of the American Statistical Association* **98**, 900–916.

Claeskens, G. and Hjort, N. L. (2008). *Model Selection and Model Averaging.* Cambridge University Press, New York.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik* **31**, 377–403.

Danilov, D. and Magnus, J. R. (2004). On the harm that ignoring pretesting can cause. *Journal of Econometrics* **122**, 27–46.

Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* **11**, 1031–1057.

Hjort, N. L. and Claeskens, G. (2003). Frequestist model average estimators. *Journal of the American Statistical Association* **98**, 879–899.

Hjort, N. L. and Claeskens, G. (2006). Focused information criteria and model averaging for the Cox hazard regression model. *Journal of the American Statistical Association* **101**, 1449–1464.

Liang, H., Härdle, W., and Carroll, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics* **27**, 1519–1535.

Magnus, J. R. and Durbin, J. (1999). Estimation of regression coefficients of interest when other regression coefficients are of no interest. *Econometrica* **67**, 639–643.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics* **15**, 661–675.

Peña, E. A., Wu, W., Piegorsch, W. W., West, R. W., and An, L. (2013). Model selection and estimation with quantal-response data in benchmark risk assessment. Tech. rep.

Schomaker, M. (2012). Shrinkage averaging estimation. *Statistical Papers* **53**, 4, 1015–1034.

Schomaker, M. and Heumann, C. (2014). Model selection and model averaging after multiple imputation. *Computational Statistics & Data Analysis* **71**, 758–770.

Schomaker, M., Wan, A. T. K., and Heumann, C. (2010). Frequentist model averaging with missing observations. *Computational Statistics & Data Analysis* **54**, 3336–3347.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)* **36**, 111–147.

Wang, H. and Zou, G. (2012). Frequentist model averaging estimation for linear errors-in-variables model. *Journal of System Science and Mathematical Science, Chinese Series* **32**, 1–14.

Wang, H., Zou, G., and Wan, A. T. K. (2012). Model averaging for varying-coefficient partially linear measurement error models. *Electronic Journal of Statistics* **6**, 1017–1039.

Wang, H., Zou, G., and Wan, A. T.-K. (2013). Adaptive lasso for varying-coefficient partially linear measurement error models. *Journal of Statistical Planning and Inference* **143**, 40–54.

You, J. and Chen, G. (2006). Estimation of a semiparametric varying-coefficient partially linear errors-in-variables model. *Journal of Multivariate Analysis* **97**, 324–341.

You, J., Zhou, Y., and Chen, G. (2006). Corrected local polynomial estimation in varying-coefficient models with measurement errors. *The Canadian Journal of Statistics* **34**, 391–410.

Zhang, X. and Liang, H. (2011). Focused information criterion and model averaging for generalized additive partial linear models. *The Annals of Statistics* **39**, 174–200.

Zhang, X., Wan, A. T. K., and Zhou, S. Z. (2012). Focused information criteria, model selection and model averaging in a tobit model with a non-zero threshold. *Journal of Business and Economic Statistics* **30**, 132–142.

Table 1: Empirical MSEs of each estimator relative to that of the full model

		$\delta^{(1)}$		$\delta^{(2)}$		$\delta^{(3)}$	
	$\sigma_u = \sigma_v =$	0.1	0.5	0.1	0.5	0.1	0.5
n=100							
μ_1	S-AIC	0.986	0.987	0.985	0.985	0.984	0.987
	AIC	1.000	1.000	0.998	1.002	0.992	1.002
	S-BIC	0.979	0.977	0.977	0.975	0.978	0.977
	BIC	0.982	0.990	0.986	0.984	0.986	0.986
	S-FIC	0.968	0.958	0.968	0.957	0.971	0.960
	FIC	0.965	0.953	0.964	0.951	0.969	0.956
μ_2	S-AIC	0.722	0.743	0.687	0.733	0.665	0.725
	AIC	0.862	0.869	0.834	0.865	0.804	0.848
	S-BIC	0.642	0.624	0.556	0.596	0.516	0.586
	BIC	0.754	0.720	0.636	0.677	0.576	0.658
	S-FIC	0.566	0.474	0.391	0.348	0.337	0.320
	FIC	0.671	0.521	0.399	0.351	0.308	0.310
n=200							
μ_1	S-AIC	0.992	0.989	0.994	0.989	0.993	0.988
	AIC	0.991	0.992	0.996	0.993	0.997	0.993
	S-BIC	0.990	0.981	0.991	0.982	0.990	0.981
	BIC	0.996	0.985	0.997	0.982	0.998	0.982
	S-FIC	0.986	0.973	0.987	0.973	0.986	0.973
	FIC	0.986	0.968	0.987	0.969	0.985	0.967
μ_2	S-AIC	0.763	0.783	0.704	0.760	0.668	0.753
	AIC	0.937	0.884	0.863	0.885	0.806	0.881
	S-BIC	0.692	0.672	0.554	0.610	0.491	0.593
	BIC	0.809	0.768	0.627	0.698	0.568	0.678
	S-FIC	0.649	0.531	0.430	0.382	0.352	0.343
	FIC	0.769	0.580	0.443	0.383	0.333	0.333

Table 2: Estimation results for CSFII data

	μ_1	μ_2	μ_3	μ_4	μ_5
	Consider the measurement errors				
S-AIC	0.443 (0.420, 0.456)	0.185 (0.164, 0.221)	0.501 (0.494, 0.515)	1.055 (1.020, 1.280)	2.399 (1.829, 2.626)
AIC	0.440 (0.422, 0.458)	0.191 (0.163, 0.218)	0.504 (0.491, 0.517)	1.043 (0.971, 1.114)	2.307 (1.910, 2.705)
S-BIC	0.443 (0.420, 0.456)	0.185 (0.164, 0.221)	0.500 (0.494, 0.515)	1.055 (1.020, 1.280)	2.403 (1.886, 2.683)
BIC	0.439 (0.421, 0.458)	0.191 (0.164, 0.219)	0.503 (0.490, 0.517)	1.012 (0.956, 1.068)	2.300 (1.905, 2.696)
S-FIC	0.444 (0.420, 0.456)	0.182 (0.164, 0.221)	0.499 (0.494, 0.515)	1.034 (1.020, 1.280)	2.585 (1.908, 2.704)
FIC	0.451 (0.436, 0.467)	0.175 (0.164, 0.185)	0.491 (0.482, 0.500)	0.960 (0.873, 1.046)	2.560 (2.373, 2.747)
Full	0.438 (0.420, 0.456)	0.193 (0.164, 0.221)	0.504 (0.494, 0.515)	1.150 (1.020, 1.280)	2.275 (1.877, 2.673)
	Ignore the measurement errors				
S-AIC	0.446 (0.428, 0.459)	0.176 (0.168, 0.191)	0.497 (0.491, 0.509)	1.068 (1.024, 1.206)	2.531 (2.224, 2.676)
AIC	0.444 (0.429, 0.46)	0.179 (0.167, 0.19)	0.500 (0.491, 0.509)	1.073 (1.037, 1.109)	2.488 (2.269, 2.707)
S-BIC	0.446 (0.428, 0.459)	0.176 (0.168, 0.191)	0.497 (0.491, 0.509)	1.068 (1.024, 1.206)	2.533 (2.248, 2.7)
BIC	0.446 (0.431, 0.462)	0.174 (0.164, 0.185)	0.496 (0.487, 0.505)	1.046 (1.031, 1.061)	2.560 (2.353, 2.767)
S-FIC	0.447 (0.428, 0.459)	0.176 (0.168, 0.191)	0.495 (0.491, 0.509)	1.046 (1.024, 1.206)	2.564 (2.25, 2.702)
FIC	0.451 (0.436, 0.467)	0.174 (0.164, 0.185)	0.492 (0.483, 0.5)	0.979 (0.9, 1.057)	2.560 (2.353, 2.767)
Full	0.444 (0.428, 0.459)	0.179 (0.168, 0.191)	0.500 (0.491, 0.509)	1.115 (1.024, 1.206)	2.473 (2.247, 2.699)

Table 3: Model selection results for CSFII data

		Consider the measurement errors	Ignore the measurement errors
AIC		x_1, x_2, x_3, x_4, x_7, x_9	x_1, x_2, x_3, x_4, x_6, x_7, x_9
BIC		x_1, x_2, x_3, x_4, x_7	x_1, x_2, x_3, x_4
FIC	μ_1	x_1, x_2, x_3, x_5, x_9	x_1, x_2, x_3, x_5, x_9
	μ_2	x_1, x_2, x_3, x_4, x_9, x_{10}, x_{11}	x_1, x_2, x_3, x_4, x_9, x_{11}
	μ_3	x_1, x_2, x_3	x_1, x_2, x_3, x_7
	μ_4	x_1, x_2, x_3, x_4, x_7, x_8, x_{10}	x_1, x_2, x_3, x_4, x_7, x_8, x_{10}
	μ_5	x_1, x_2, x_3, x_4	x_1, x_2, x_3, x_4