SFB 823

Discussion Paper

Choice is suffering: A focused information criterion for model selection

Peter Behl, Holger Dette, Manuel Frondel, Harald Tauchmann

Nr. 42/2010

SFB 823

Choice is Suffering: A Focused Information Criterion for Model Selection

Peter Behl, **Holger Dette**, Ruhr University Bochum, **Manuel Frondel**, Ruhr University Bochum and RWI, **Harald Tauchmann**, RWI

Abstract. In contrast to conventional measures, the Focused Information Criterion (FIC) allows the purpose-specific selection of models, thereby reflecting the idea that one kind of model might be appropriate for inferences on a parameter of interest, but not for another. Ever since its introduction, the FIC has been increasingly applied in the realm of statistics, but this concept appears to be virtually unknown in the economic literature. Using a straightforward analytical example, this paper illustrates the FIC and its usefulness in economic applications.

JEL classification: C3, D2.

Key words: AKAIKE Information Criterion, SCHWARZ Information Criterion, Translog Cost Function.

Correspondence: Prof. Dr. Manuel Frondel, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Hohenzollernstr. 1-3, D-45128 Essen. frondel@rwi-essen.de.

Acknowledgements: We are greatly indebted to Christoph M. Schmidt and Colin Vance for their invaluable comments and suggestions. This work has been supported in part by the Collaborative Research Center "Statistical Modelling of Nonlinear Dynamic Processes" (SFB 823) of the German Research Foundation (DFG).

1 Introduction

Selecting an adequate model is key for any empirical analysis. Numerous methods for model choice and validation have been suggested in the literature. Well-known approaches to model selection include the usage information criteria, such as AKAIKE's (1970) AIC and SCHWARZ' (1978) SIC.[1] Alternatively, DETTE (1999), DETTE, PODOLSKIJ and VETTER (2006), or PODOLSKIJ and DETTE (2008) propose, among many others, goodness-of-fit tests. Common to all these tests, measures, and criteria is the idea that they provide us with a single 'best' model, regardless of the purpose of inference. Deviating from this conventional avenue, CLAESKENS and HJORT (2003) have conceived the Focused Information Criterion (FIC) to allow various models to be selected for different purposes.

This approach reflects the view that one kind of model might be appropriate for inferences on, say, the cross-price elasticity of capital and labor, whereas a different sort of model may be preferable for the estimation of another parameter. Ever since its introduction, the FIC has been increasingly applied in the realm of statistics, but the concept appears to be virtually unknown in the economic literature. Using the classical example of the choice among COBB-DOUGLAS- and translog models for didactic purposes, this paper illustrates the concept and usefulness of the FIC, focusing on the substitutability of capital and labor.

The following Section 2 describes the classical example and the focus parameter. Section 3 explains the core of the FIC, the information matrix, and calculates it for our analytical example. In Section 4, we apply the FIC to the model selection problem presented in Section 2. The last section summarizes.

[1] According to KENNEDY (2003:117), AIC tends to select models that are over-parameterized, whereas SIC, which is also termed Bayesian Information Criterion (BIC), tends to pick up the true model if this is among the choices. The SIC is considered by most researchers to be the best criterion, as it has performed well in Monte-Carlo studies.

2 A Classical Example

We use the frequently employed translog cost function approach – see e. g. FRONDEL and SCHMIDT (2002, 2003) for surveys – including here merely two inputs, capital (K) and labor (L), where p_K and p_L denote the respective prices:

$$\begin{aligned}\log C(p_K, p_L) &= \beta_0 + \beta_K \log p_K + \beta_L \log p_L \\ &+ \frac{1}{2}\beta_{KK} \log p_K \log p_K + \beta_{KL} \log p_K \log p_L \\ &+ \frac{1}{2}\beta_{LL} \log p_L \log p_L.\end{aligned} \tag{1}$$

This approach reduces to the COBB-DOUGLAS function if the second-order coefficients β_{KK}, β_{LL}, and β_{KL} vanish:

$$H_0 : \beta_{KK} = \beta_{LL} = \beta_{KL} = 0. \tag{2}$$

Given empirical data on input prices, as well as on cost shares of capital (s_K) and labor (s_L), an efficient procedure to obtain coefficient estimates is via a cost share system (BERNDT, 1996:470):

$$\begin{aligned} s_K &= \beta_K + \beta_{KK} \log p_K + \beta_{KL} \log p_L, \\ s_L &= \beta_L + \beta_{KL} \log p_K + \beta_{LL} \log p_L, \end{aligned} \tag{3}$$

which results from the logarithmic differentiation of translog function (1) with respect to p_K and p_L, respectively, as e. g. $\frac{\partial \log C}{\partial \log p_K} = \frac{p_K}{C}\frac{\partial C}{\partial p_K} = \frac{p_K x_K}{C} = s_K$, where according to SHEPARD's Lemma $\frac{\partial C}{\partial p_K} = x_K$.

In this two-factor case, cost share system (3) degenerates to a single cost share equation:

$$s_K = \beta_K + \beta_{KK} \log(p_K/p_L), \tag{4}$$

as both cost shares add to unity, $s_K + s_L = 1$, thereby implying the following restrictions that are already incorporated in (4):

$$1 = \beta_K + \beta_L, \tag{5}$$

$$0 = \beta_{KK} + \beta_{KL}, \quad (6)$$

$$0 = \beta_{KL} + \beta_{LL}. \quad (7)$$

On the basis of (4), the classical procedure of selecting either of the two specifications involves testing whether β_{KK} equals zero:

$$H_0 : \beta_{KK} = 0. \quad (8)$$

Alternatively, using the FIC for model selection requires determining a parameter of interest μ, which is typically a function of the model coefficients. As in many empirical labor market studies, we focus here on the capital elasticity with respect to wages, η_{Kp_L}, which for the translog cost function (1) is given by (see e. g. FRONDEL and SCHMIDT (2006:188))

$$\mu(\beta_K, \beta_L, \beta_{KK}, \beta_{KL}, \beta_{LL}, \sigma) := \eta_{Kp_L} = \frac{\beta_{KL}}{s_K} + s_L. \quad (9)$$

This expression degenerates to $\eta_{Kp_L} = s_L$ for the COBB-DOUGLAS function, as can be seen from hypothesis (8) and restriction (6).

3 Information Measures and Matrices

Using the abbreviation $X := \log(p_K/p_L)$ and re-notating s_K by $Y := s_K$, the stochastic version of the more general specification (4) reads

$$Y = \beta_K + \beta_{KK}X + \varepsilon, \quad (10)$$

where ε denotes the error term, whose variance structure is assumed to be homoscedastic: Var(ε) = σ^2. In line with CLAESKENS and HJORT (2003:91), specification (10) is called here *full* model. Relative to the so-called narrow model, also referred to as the *null* model, the single parameter $\gamma := \beta_{KK}$ completes the full model. For clarity, the parameters estimated from the full model are designated by $\boldsymbol{\theta}^{full} := (\beta_K^{full}, \sigma^{full}, \gamma^{full})^T$, where $\gamma^{full} = \beta_{KK}^{full}$, whereas those of the null model are denoted by $\boldsymbol{\theta}^0 := (\beta_K^0, \sigma^0, \gamma^0)^T$. Corresponding to (8), γ^0 equals zero: $\gamma^0 = 0$.

For comparing competing parametric models on the basis of an n-dimensional sample that provides observations $(x_1, ..., x_n)$ and $(y_1, ..., y_n)$ on X and Y, respectively, applying FIC requires the maximum likelihood (ML) estimation method (CLAESKENS, HJORT, 2003:91) and the calculation of a (p+q) $\times$ (p+q) information matrix, where p refers to the number of parameters estimated in the null model and q designates the number of parameters that exclusively belong to the full model. In our example, $p = 2$ and $q = 1$, that is, I^{full} is a 3 $\times$ 3 matrix and I_{00} is a 2 $\times$ 2 matrix, whereas I_{11} is a scalar:

$$I^{full} := \begin{pmatrix} I_{00} & I_{01} \\ I_{10} & I_{11} \end{pmatrix} = E \begin{bmatrix} \left(\frac{\partial \log L}{\partial \beta_K}\right)^2 & \frac{\partial \log L}{\partial \beta_K}\frac{\partial \log L}{\partial \sigma} & \frac{\partial \log L}{\partial \beta_K}\frac{\partial \log L}{\partial \gamma} \\ \frac{\partial \log L}{\partial \sigma}\frac{\partial \log L}{\partial \beta_K} & \left(\frac{\partial \log L}{\partial \sigma}\right)^2 & \frac{\partial \log L}{\partial \sigma}\frac{\partial \log L}{\partial \gamma} \\ \frac{\partial \log L}{\partial \gamma}\frac{\partial \log L}{\partial \beta_K} & \frac{\partial \log L}{\partial \gamma}\frac{\partial \log L}{\partial \sigma} & \left(\frac{\partial \log L}{\partial \gamma}\right)^2 \end{bmatrix}. \tag{11}$$

The entries of I^{full} are based on FISHER's well-known information measure. When focusing on parameter γ, the respective entry is given by

$$I_{11} = E[\left(\frac{\partial \log L(\beta_K, \sigma, \gamma, X)}{\partial \gamma}\right)^2] = E[\left(\frac{\partial L(\beta_K, \sigma, \gamma, X)}{\partial \gamma} / L(\beta_K, \sigma, \gamma, X)\right)^2].$$

FISHER's information measure helps to discriminate between two parameter values γ_1 and γ_2 on the basis of the likelihood $L(\beta_K, \sigma, \gamma, X)$. Intuitively, the larger the difference $L(\beta_K, \sigma, \gamma_1, X) - L(\beta_K, \sigma, \gamma_2, X)$, the more easy it is to discriminate between γ_1 and γ_2. FISHER's measure captures this difference by the partial derivative of the likelihood, $\partial \log L/\partial\gamma$, relative to the likelihood L. This ratio is squared in order to account for positive and negative relative differences alike. Finally, to obtain a global measure that is independent of individual samples, expectations are built.

To determine the entries of information matrix I^{full}, we assume normality of the error term: $\varepsilon \sim N(0, \sigma^2)$. The log-likelihood of ε then reads

$$\log L = -\log\sqrt{2\pi} - \log\sigma - \frac{1}{2}\left(\frac{Y - \beta_K - \beta_{KK}X}{\sigma}\right)^2. \tag{12}$$

Given this log-likelihood, we get

$$\frac{\partial \log L}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}^0} = \begin{pmatrix} \frac{\partial \log L}{\partial \beta_K} \\ \frac{\partial \log L}{\partial \sigma} \\ \frac{\partial \log L}{\partial \beta_{KK}} \end{pmatrix}_{|\boldsymbol{\theta}^0} = \begin{pmatrix} \frac{Y-\beta_K-\beta_{KK}\cdot X}{\sigma} \cdot \frac{1}{\sigma} \\ -\frac{1}{\sigma} + \frac{(Y-\beta_K-\beta_{KK}\cdot X)^2}{\sigma^3} \\ \frac{X(Y-\beta_K-\beta_{KK}\cdot X)}{\sigma} \cdot \frac{1}{\sigma} \end{pmatrix}_{|\boldsymbol{\theta}^0} = \begin{pmatrix} \frac{\varepsilon^0}{\sigma^0} \\ \frac{(\varepsilon^0)^2-1}{\sigma^0} \\ \frac{X\varepsilon^0}{\sigma^0} \end{pmatrix}, \quad (13)$$

where $\boldsymbol{\theta} := (\beta_K, \sigma, \gamma = \beta_{KK})^T$, $\boldsymbol{\theta}^0 := (\beta_K^0, \sigma^0, \gamma^0 = 0)^T$, and $\varepsilon^0 := \frac{Y-\beta_K^0}{\sigma^0} \sim N(0,1)$.

Using vector $\frac{\partial \log L}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}^0}$ as given by (13) and evaluating the information matrix I^{full} at $\boldsymbol{\theta}^0$, the common anchor of both models, yields

$$\begin{aligned} I^{full}|_{\boldsymbol{\theta}^0} &= E[\left(\frac{\partial \log L}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}^0}\right) \cdot \left(\frac{\partial \log L}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}^0}\right)^T] = \begin{bmatrix} E[(\frac{\varepsilon^0}{\sigma^0})^2] & E[\frac{\varepsilon^0}{\sigma^0}\frac{(\varepsilon^0)^2-1}{\sigma^0}] & E[\frac{\varepsilon^0}{\sigma^0}\frac{X\varepsilon^0}{\sigma^0}] \\ E[\frac{\varepsilon^0}{\sigma^0}\frac{(\varepsilon^0)^2-1}{\sigma^0}] & E[(\frac{(\varepsilon^0)^2-1}{\sigma^0})^2] & E[\frac{(\varepsilon^0)^2-1}{\sigma^0}\frac{X\varepsilon^0}{\sigma^0}] \\ E[\frac{\varepsilon^0}{\sigma^0}\frac{X\varepsilon^0}{\sigma^0}] & E[\frac{(\varepsilon^0)^2-1}{\sigma^0}\frac{X\varepsilon^0}{\sigma^0}] & E[(\frac{X\varepsilon^0}{\sigma^0})^2] \end{bmatrix} \\ &= \frac{1}{(\sigma^0)^2} \begin{bmatrix} 1 & 0 & X \\ 0 & 2 & 0 \\ X & 0 & X^2 \end{bmatrix}, \end{aligned} \quad (14)$$

as $E[(\varepsilon^0)^2]$ = Var $(\varepsilon^0) = 1$, $E[\varepsilon^0] = 0 = E[(\varepsilon^0)^3]$, and $E[(\varepsilon^0)^4] = 3$.

Employing the methods of moments provides an estimate of $I^{full}|_{\boldsymbol{\theta}^0}$:

$$\hat{I}^{full}|_{\boldsymbol{\theta}^0} = \begin{pmatrix} \hat{I}_{00} & \hat{I}_{01} \\ \hat{I}_{10} & \hat{I}_{11} \end{pmatrix} = \frac{1}{(\widehat{\sigma^0})^2} \begin{pmatrix} 1 & 0 & \bar{x} \\ 0 & 2 & 0 \\ \bar{x} & 0 & \overline{x^2} \end{pmatrix}, \quad (15)$$

with $\bar{x} := (x_1 + ... + x_n)/n$, $\overline{x^2} := (x_1^2 + ... + x_n^2)/n$, and $(\widehat{\sigma^0})^2$ being the ML-estimate of $(\sigma^0)^2$.

4 One-Dimensional FIC

In our one-dimensional illustrative example, in which both models differ in merely the single coefficient $\gamma = \beta_{KK}$, the FIC reduces for the null model to

(CLAESKENS, HJORT, 2003:907):[2]

$$FIC^0 = \omega^2 D^2, \tag{16}$$

where

$$D := \sqrt{n}(\gamma^{full} - \gamma^0) = \sqrt{n}\gamma^{full}, \tag{17}$$

as $\gamma^0 = 0$, and

$$\omega := I_{10} I_{00}^{-1} \frac{\partial \mu}{\partial \boldsymbol{\xi}}|_{\boldsymbol{\theta}^0} - \frac{\partial \mu}{\partial \gamma}|_{\boldsymbol{\theta}^0}, \tag{18}$$

with $\boldsymbol{\xi} := (\beta_K, \sigma)^T$. For the full model, the FIC is given by

$$FIC^{full} = 2\omega^2 K, \tag{19}$$

with

$$K := (I_{11} - I_{10} I_{00}^{-1} I_{01})^{-1}. \tag{20}$$

Using $\hat{I}^{full}|_{\boldsymbol{\theta}^0}$ from (15), we get a familiar estimate of K:

$$\begin{aligned} \hat{K} &= (\hat{I}_{11} - \hat{I}_{10} \hat{I}_{00}^{-1} \hat{I}_{11})^{-1} = [\frac{\overline{x^2}}{(\widehat{\sigma^0})^2} - (\frac{\bar{x}}{(\widehat{\sigma^0})^2}, 0) \begin{pmatrix} (\widehat{\sigma^0})^2 & 0 \\ 0 & (\widehat{\sigma^0})^2/2 \end{pmatrix} \begin{pmatrix} \frac{\bar{x}}{(\widehat{\sigma^0})^2} \\ 0 \end{pmatrix}]^{-1} \\ &= [\frac{\overline{x^2}}{(\widehat{\sigma^0})^2} - (\frac{\bar{x}}{(\widehat{\sigma^0})^2}, 0) \begin{pmatrix} \bar{x} \\ 0 \end{pmatrix}]^{-1} = \frac{(\widehat{\sigma^0})^2}{\overline{x^2} - (\bar{x})^2}, \end{aligned} \tag{21}$$

which is proportional to the variance of the ML-estimate $\hat{\gamma}^{full} = \hat{\beta}_{KK}^{full}$. Note that $\hat{\beta}_{KK}^{full}$ is the essential ingredient of the estimate $\hat{D} = \sqrt{n}\hat{\beta}_{KK}^{full}$ of bias $D = \sqrt{n}(\gamma^{full} - \gamma^0)$. In short, irrespective of the concrete value of the common term ω, comparing FIC^0 and FIC^{full} in fact reflects the trade-off of bias D versus estimation variability given by K.

While – as a rule of guidance – the (sub-)model with the smallest estimate of FIC is chosen, for the nontrivial case in which $\omega \neq 0$, the narrow model is preferred by the FIC over the full model if $FIC^0 = \omega^2 D^2 < 2\omega^2 K = FIC^{full}$ or,

[2] Ultimately, it will turn out that the application of the FIC becomes irrelevant in this one-dimensional case.

equivalently, if $D^2/K < 2$ (CLAESKENS, HJORT, 2003:907). In our example, this decision is based on the following estimate of a $\chi^2(1)$-distributed test statistic:

$$\frac{\hat{D}^2}{\hat{K}} = \frac{(\hat{\beta}_{KK}^{full})^2}{\frac{(\widehat{\sigma^0})^2}{n(\overline{x^2}-(\bar{x})^2)}},$$

where $\frac{(\widehat{\sigma^0})^2}{n(\overline{x^2}-(\bar{x})^2)}$ is the variance of $\hat{\beta}_{KK}^{full}$ and the significance level results from $Pr(\chi^2(1) \geq 2) = 0.157$.

Although in our example the decision on whether to prefer the null or the full model does not depend upon the choice of the focus parameter μ at all, for illustrative purposes, we nonetheless calculate the FIC both for our preferred focus parameter

$$\mu = \eta_{Kp_L} = \frac{\beta_{KL}}{s_K} + s_L = \frac{-\beta_{KK}}{\beta_K + \beta_{KK}\log X} + 1 - (\beta_K + \beta_{KK}\log X), \qquad (22)$$

and, alternatively, for $\mu = \beta_{KK} = \gamma$, for which $\frac{\partial\mu}{\partial\gamma} = 1$, $\frac{\partial\mu}{\partial\boldsymbol{\xi}} = (0,0)^T$, and hence $\omega = -1$, so that $FIC^0 = D^2$ and $FIC^{full} = 2K$.

In contrast, for $\mu = \eta_{Kp_L}$, we obtain from expression (22)

$$\begin{aligned}
\frac{\partial\mu}{\partial\gamma}|_{\boldsymbol{\theta}^0} &= \frac{\partial\mu}{\partial\beta_{KK}}|_{\boldsymbol{\theta}^0} = (\frac{-\beta_K}{(\beta_K + \beta_{KK}\log X)^2} - X)|_{\boldsymbol{\theta}^0} = -\frac{1}{\beta_K^0} - X, \\
\frac{\partial\mu}{\partial\boldsymbol{\xi}}|_{\boldsymbol{\theta}^0} &= \begin{pmatrix} \frac{\beta_{KK}}{\beta_K+\beta_{KK}\log X} - 1, \\ 0 \end{pmatrix}|_{\boldsymbol{\theta}^0} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.
\end{aligned}$$

Using these derivatives and definition (18), for $X = \bar{x}$ the estimate of ω reads

$$\hat{\omega} = \hat{I}_{10}\hat{I}_{00}^{-1}\widehat{\frac{\partial\mu}{\partial\boldsymbol{\xi}}}|_{\boldsymbol{\theta}^0} - \widehat{\frac{\partial\mu}{\partial\gamma}}|_{\boldsymbol{\theta}^0} = (\frac{\bar{x}}{(\widehat{\sigma^0})^2}, 0)\begin{pmatrix} (\widehat{\sigma^0})^2 & 0 \\ 0 & (\widehat{\sigma^0})^2/2 \end{pmatrix}\begin{pmatrix} -1 \\ 0 \end{pmatrix} + \frac{1}{\hat{\beta}_K^0} + \bar{x} = \frac{1}{\hat{\beta}_K^0}.$$

In sum, $\widehat{FIC}^{full} = 2\hat{\omega}^2\hat{K} = \frac{2}{(\hat{\beta}_K^0)^2}\frac{(\widehat{\sigma^0})^2}{\overline{x^2}-(\bar{x})^2}$, which in accord with $(\widehat{\sigma^0})^2$ should be close to zero if translog function (1) is the true model. Similarly intuitive is that $\widehat{FIC}^0 = \hat{\omega}\hat{D} = n(\frac{\hat{\beta}_{KK}^{full}}{\hat{\beta}_K^0})^2$ should be small or even vanish if COBB-DOUGLAS is the true model and, hence, β_{KK}^{full} is close to, or even equals, zero.

It bears noting that $\hat{\omega}$ generally depends upon the concrete value $X = x$:

$$\hat{\omega} = -\bar{x} + \frac{1}{\hat{\beta}_K^0} + x, \qquad (23)$$

so that the FIC also critically hinges on the individual value $X = x$. As a consequence, it may well be the case that with this criterion the full model might be preferred for some, but not for all x. In contrast to other measures, such as AIC, the FIC therefore does not provide for a unanimous model recommendation across the whole range of values of the conditional variables.[3]

More generally, in the q-dimensional case in which the models under scrutiny may differ in q parameters $\gamma_1, ..., \gamma_q$, the FIC is given by

$$FIC := (\sum_{j=1}^{q} \omega_j D_j 1(\gamma_j = \gamma_j^0))^2 + 2\sum_{j=1}^{q} \omega_j^2 K_j 1(\gamma_j \neq \gamma_j^0), \qquad (24)$$

if K is diagonal with entries K_j and where 1(.) denotes the indicator function. Note that for $q = 1$ definition (24) specializes to either (16) if $\gamma = \gamma^0$ or (19) if $\gamma \neq \gamma^0$, with $\omega_1 = \omega$ being a common ingredient. For $q > 1$, the factors $\omega_1, ..., \omega_q$, which vary with the focus parameter μ, generally differ from each other. Thus, as opposed to the one-dimensional case illustrated here, different models may be preferred by the FIC in the multi-dimensional case, depending upon the concrete choice of focus parameter μ.

5 Summary

Econometric studies on factor substitution frequently stress the importance of choosing the right model for correctly describing the true technology of production (e.g. CONSIDINE, 1989). Typically, this choice focuses on a few well-established functional forms, such as Leontief, linear-logit, and, often, translog. In seeking the right functional form, however, one might overlook that any parametric model represents a highly stylized description of the real production process. As a consequence, none of these functional forms can claim to be the true model, albeit they may capture certain features of reality reasonably well. Rather than looking for the ultimately true model, an alternative avenue is to look for

[3] Alternatively, one might use a weighted version of the FIC (see CLAESKENS, HJORT, 2008).

that functional form that is most appropriate for answering a specific research question, such as the substitution relationship of energy and capital.

This is precisely the core of the concept of the Focused Information Criterion (FIC), developed by CLAESKENS and HJORT (2003) to allow for purpose-specific model selection. Using a one-dimensional analytical example, this paper has illustrated this concept. Its underlying idea is to study perturbations of a parametric model that rests on the parameters $\boldsymbol{\gamma}^0 := (\gamma_1^0, ..., \gamma_q^0)^T$ as a point of departure, with $\boldsymbol{\gamma}^0$ being known. A variety of models may then be considered that depart from $\boldsymbol{\gamma}^0$ in some or all of q directions: $\boldsymbol{\gamma} \neq \boldsymbol{\gamma}^0$.

On the basis of the maximum-likelihood estimates for the parameters of the altogether 2^q (sub-)models, that model for which the FIC is minimal for a given focus parameter of choice $\mu = \mu(\boldsymbol{\gamma})$ will be selected, a selection procedure that – except for the one-dimensional case $q = 1$ – critically hinges on the choice of the focus parameter μ. In contrast, classical selection criteria are not related to the purpose of inference. In addition to this feature, the FIC contrasts with other model selection measures, such as the AKAIKE and SCHWARZ criteria, in that it is not a global criterion that recommends a single, most preferred model irrespective of the values of the covariates. Rather, it is a local criterion that may indicate the appropriateness of various models, depending upon the vicinity of the conditioning variables.

References

AKAIKE, H. (1970) Statistical Predictor Identification. *Annals of the Institute of Statistical Mathematics* 22, 203-217.

BERNDT, E. R. (1996) *The Practice of Econometrics - Classic and Contemporary*. Addison-Wesley Publishing Company.

CLAESKENS, G. , HJORT, N. L. (2003) The Focused Information Criterion. *Journal of the American Statistical Association* 98, 900-916.

CLAESKENS, G. , HJORT, N. L. (2008) Minimising average risk in regression models. *Econometric Theory* 24, 493-527.

CONSIDINE, T. J. (1989) Separability, Functional form and Regulatory Policy in Models of Inter-fuel Substitution. *Energy Economics* 11, 82-94.

DETTE, H. (1999) A Consistent Test for the Functional Form of a Regression based on a Difference of Variance Estimators. *Annals of Statistics*, 27, 1012-1040.

DETTE, H., PODOLSKIJ, M. , VETTER, M. (2006) Estimation of Integrated Volatility in Continuous Time Financial Models With Applications to Goodness-of-Fit Testing. *Scandinavian Journal of Statistics*, 33, 259-278.

FRONDEL, M., SCHMIDT, C. M. (2006) The Empirical Assessment of Technology Differences: Comparing the Comparable. *The Review of Economics and Statistics*, Vol. 88 (1), 186-192.

FRONDEL, M., SCHMIDT, C. M. (2003) Rejecting Capital-Skill Complementarity at All Costs? *Economics Letters*, Vol. 80, Issue 1, 15-21.

FRONDEL M. , SCHMIDT, C. M. (2002) The Capital-Energy Controversy: An Artifact of Cost Shares. *The Energy Journal*, Vol. 23, Issue 3, 53-79.

KENNEDY, P. (2003) A Guide to Econometrics. Fifth edition. Blackwell Publishing.

PODOLSKIJ, M. DETTE, H. (2008) Testing the Parametric Form of the Volatility in Continuous Time Diffusion Models - a Stochastic Process Approach. *Journal of Econometrics* 143, 56-73.

SCHWARZ, G. (1978) Estimating the Dimension of a Model. *Annals of Statistics* 6, 461464.