Being Focused: When the Purpose of Inference Matters for Model Selection

Peter Behl, **Holger Dette**, Ruhr University Bochum (RUB), **Manuel Frondel**, Ruhr University Bochum (RUB), Ruhr Graduate School in Economics (RGS Econ), and Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), **Harald Tauchmann**, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)

Abstract. In contrast to conventional model selection criteria, the Focused Information Criterion (FIC) allows for purpose-specific choice of models. This accommodates the idea that one kind of model might be highly appropriate for inferences on a particular parameter, but not for another. Ever since its development, 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 classical example and data for 35 U. S. industry sectors (1960-2005), this paper provides for an illustration of the FIC and a demonstration of its usefulness in empirical applications.

JEL classification: C3, D2.

Key words: Information Criteria, Translog Cost Function, Cross-Price Elasticities.

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 by the Collaborative Research Center "Statistical Modelling of Nonlinear Dynamic Processes" (SFB 823) of the German Research Foundation (DFG), within the framework of project A3, "Dynamic Technology Modelling", and project C1, "Model Choice and Dynamic Dependence Structures". The work of Peter Behl has been supported by a doctoral scholarship of the Hanns-Seidel-Foundation.

1 Introduction

There is an impressive bulk of empirical studies that aim at measuring the ease of substitution between production factors (for surveys, see e. g. KINTIS, PANAS, 1989, or FRONDEL, SCHMIDT, 2003), with a growing emphasis on the substitution relationships of energy with respect to other inputs (see e. g. APOSTOLAKIS, 1990, or FRONDEL, SCHMIDT, 2002, 2004). Common to the overwhelming majority of these studies is that the substitution parameters of interest are gleaned from a single 'best' model that is estimated on the basis of the empirical data at hand, but regardless of the purpose of inference.

To this end, model selection methods include the usage of information criteria, such as AKAIKE's (1974) AIC and SCHWARZ' (1978) SIC. Alternatively, DETTE (1999), DETTE, PODOLSKIJ and VETTER (2006), or PODOLSKIJ and DETTE (2008) propose, among many others, goodness-of-fit tests. Typically, the selection of the most appropriate model focuses on a few well-established functional forms, such as Generalized LEONTIEF, COBB-DOUGLAS, and, most often, the Translog model. In seeking the right functional form, however, one might ignore 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. Most importantly, depending on the facet of reality that is the focus of the analysis, divergent specifications might approximate different facets in an optimal way.

Recognizing this argument, CLAESKENS and HJORT (2003) deviated from the conventional avenue and conceived the Focused Information Criterion (FIC) to allow various models to be selected for different purposes. In the illustrative case of the estimation of the degree of substitutability of capital and energy and of labor and energy, one kind of model might be highly appropriate for inferences on, say, the cross-price elasticity of capital with respect to energy prices, whereas a different sort of model may be preferable for the estimation of another parameter, such as the cross-price elasticity of labor with respect to energy prices. Because of its usefulness in balancing modeling bias against estimation variability, the FIC has been increasingly applied in the realm of statistics, but, with BEHL et al. (2010) being an exception, this concept appears to be virtually unknown in the economic literature.

Using the classical example of the choice among COBB-DOUGLAS and Translog models and data for 35 U.S. industry sectors in the time period spanning 1960 to 2005, this article illustrates the concept and usefulness of the FIC, thereby demonstrating that the selection of a model type critically depends on the purpose of inference. From our three-factor example, it will become evident that this choice is highly dependent on the focus parameter μ, that is, whether the cross-price elasticity for either labor or capital demand with respect to energy prices is the primary aim of the analysis.

The general idea underlying the FIC, which ultimately results from estimating the mean squared error of the modeling bias (CLAESKENS, HJORT, 2003:902), is to study perturbations of a parametric model, with the known parameter vector $\gamma^0 := (\gamma_1^0, ..., \gamma_q^0)^T$ as the point of departure, which in our example will be set to zero without any loss of generality: $\gamma^0 = \mathbf{0}$. A variety of models may then be considered that depart from γ^0 in some or all of q directions: $\gamma \neq \gamma^0$. On the basis of parameter estimates of the altogether 2^q sub-models, that candidate model for which the FIC is minimal for a given focus parameter of choice $\mu = \mu(\gamma)$ will be selected. By minimizing the FIC, one captures the trade-off between modeling bias, which, by definition, is zero for the most general model for which $\gamma_1 \neq \gamma_1^0, ..., \gamma_q \neq \gamma_q^0$, and relative estimation variability, which, by definition, is zero for the most restricted model for which $\gamma_1 = \gamma_1^0, ..., \gamma_q = \gamma_q^0$. In our example, we confine ourselves to these two polar model specifications, the most general and the most restricted model, rather than estimating all of the 2^q model specifications.

The following Section 2 presents our example and derives the analytical expressions needed for the model selection among COBB-DOUGLAS and Translog on the basis of the FIC. Section 3 provides for a concise introduction into the concept of the FIC, followed by the presentation of the empirical example in Section 4. The last section summarizes and concludes.

2 Analytical Example

To illustrate the concept of the FIC on the basis of a straightforward example that is – for the sake of simplicity – restricted to the case of three production factors, we employ the dual approach (BERNDT, 1996), in which a system of cost share equations is derived from the underlying cost function via SHEPARD's lemma. For a COBB-DOUGLAS cost function, cost shares are well-known to be independent from factor prices:

$$s_K = \beta_K + \nu_K, \qquad s_L = \beta_L + \nu_L, \tag{1}$$

where s_K and s_L denote the cost shares of capital K and labor L, respectively, β_L and β_K are parameters to be estimated, and ν_L and ν_K are random errors for which we assume joint normality, as Maximum Likelihood (ML) is CLAESKENS' and HJORT's (2003) estimation method of choice.

Adding a third equation for the cost share of energy E to system (1) would be superfluous, as the cost shares sum up to unity: $s_K + s_L + s_E = 1$. Implicitly, this property yields the restrictions $\beta_K + \beta_L + \beta_E = 1$ and $\nu_K + \nu_L + \nu_E = 0$, so that an estimate of $\beta_E = 1 - \beta_L - \beta_K$ can be obtained from the estimates of β_L and β_K.

For the same reason, it suffices to estimate the following two-equations system for the Translog cost function:

$$s_K = \beta_K + \beta_{KK}\, p_K + \beta_{KE}\, p_E + \nu_K, \quad s_L = \beta_L + \beta_{LK}\, p_K + \beta_{LE}\, p_E + \nu_L, \tag{2}$$

where p_K and p_E denote the logged relative factor prices $\log(\tilde{p}_K/\tilde{p}_L)$ and $\log(\tilde{p}_E/\tilde{p}_L)$, respectively, with labor being chosen as the numeraire. If $\beta_{KK} = \beta_{KE} = \beta_{LK} = \beta_{LE} = 0$, the Translog specification degenerates to COBB-DOUGLAS.

Adopting the terminology of CLAESKENS and HJORT (2003), the COBB-DOUGLAS specification (1) is called the null model. For this specification, also referred to as the narrow model, the vector $\boldsymbol{\xi}$ of parameters that are subject to estimation comprises four elements:

$$\boldsymbol{\xi} := (\beta_K, \beta_L, \sigma_K, \sigma_L)^T, \tag{3}$$

where T indicates the transposition of a vector and σ_K and σ_L designate the standard deviations of ν_L and ν_K, respectively. The vector of parameters that are additionally included in the Translog model, which is called the full model, reads:

$$\gamma := (\beta_{KK}, \beta_{KE}, \beta_{LK}, \beta_{LE}, \rho_{KL})^T, \tag{4}$$

where ρ_{KL} stands for the correlation of the error terms: $\rho_{KL} := Corr(\nu_K, \nu_L)$. For clarity, the parameters estimated from the null model are denoted by $\boldsymbol{\theta}^0 := (\boldsymbol{\xi}^0, \gamma^0)^T$, with $\boldsymbol{\xi}^0 := (\beta_K^0, \beta_L^0, \sigma_K^0, \sigma_L^0)^T$ and $\gamma^0 = \mathbf{0}$ if we additionally assume $\rho_{KL} = 0$ for the COBB-DOUGLAS case.

In contrast to conventional selection criteria, using the FIC for model selection orients towards one or more measures of interest, called here focus parameters and designated by μ, which are typically a function of the model coefficients: $\mu = \mu(\boldsymbol{\xi}, \gamma)$. As our focus is on the substitutability of energy by both labor and capital, we choose the cross-price elasticities of capital and labor demand, both with respect to the price of energy, as focus parameters. For the Translog model, these substitution elasticities are given by (see e. g. FRONDEL, SCHMIDT, 2006):

$$\eta_{K_{p_E}} = \frac{\beta_{KE}}{s_K} + s_E = \frac{\beta_{KE}}{s_K} + 1 - s_K - s_L, \tag{5}$$

$$\eta_{L_{p_E}} = \frac{\beta_{LE}}{s_L} + s_E = \frac{\beta_{LE}}{s_L} + 1 - s_K - s_L, \tag{6}$$

where according to system (2) the cost shares of capital and labor itself depend on coefficients such as β_K, β_{KK}, etc. For the special case of COBB-DOUGLAS, both elasticities degenerate to the same entity, the cost share of energy: s_E.

As we will see in the subsequent section, the dependence of the FIC on a focus measure μ – here the elasticities $\eta_{L_{p_E}}$ and $\eta_{K_{p_E}}$ – is given by the vectors of partial derivatives of such measures with respect to both the coefficients belonging to the null model, $\boldsymbol{\xi}$, and those that exclusively belong to the full model, $\boldsymbol{\gamma}$. For $\mu = \mu(\boldsymbol{\xi}, \boldsymbol{\gamma}) = \eta_{L_{p_E}}$, for instance, the partial derivatives are given by:

$$\frac{\partial \mu}{\partial \boldsymbol{\xi}} = \left(\frac{\partial \mu}{\partial \beta_K}, \frac{\partial \mu}{\partial \beta_L}, \frac{\partial \mu}{\partial \sigma_K}, \frac{\partial \mu}{\partial \sigma_L} \right)^T = (-1, -\frac{\beta_{LE}}{s_L^2} - 1, 0, 0)^T \tag{7}$$

and

$$\begin{aligned} \frac{\partial \mu}{\partial \boldsymbol{\gamma}} &= \left(\frac{\partial \mu}{\partial \beta_{KK}}, \frac{\partial \mu}{\partial \beta_{KE}}, \frac{\partial \mu}{\partial \beta_{LK}}, \frac{\partial \mu}{\partial \beta_{LE}}, \frac{\partial \mu}{\partial \rho_{KL}} \right)^T \\ &= (-p_K, -p_E, -\frac{\beta_{LE}}{s_L^2} p_K - p_K, \frac{1}{s_L} - \frac{\beta_{LE}}{s_L^2} p_E - p_E, 0)^T. \end{aligned} \tag{8}$$

Evaluating these derivatives at $\boldsymbol{\theta}^0 = (\boldsymbol{\xi}^0, \gamma^0)^T = (\boldsymbol{\xi}^0, \mathbf{0})^T$ yields the vectors

$$\frac{\partial \eta_{L_{p_E}}}{\partial \boldsymbol{\xi}}|_{\boldsymbol{\theta}^0} = (-1, -1, 0, 0)^T, \ \frac{\partial \eta_{L_{p_E}}}{\partial \boldsymbol{\gamma}}|_{\boldsymbol{\theta}^0} = (-p_K, -p_E, -p_K, \frac{1}{\beta_L^0} - p_E, 0)^T, \tag{9}$$

which will be required for the calculation of the FIC in our empirical example. Similarly, for focus parameter $\mu = \eta_{K_{p_E}}$, the partial derivatives read:

$$\frac{\partial \eta_{K_{p_E}}}{\partial \boldsymbol{\xi}}|_{\boldsymbol{\theta}^0} = (-1, -1, 0, 0)^T, \ \frac{\partial \eta_{K_{p_E}}}{\partial \boldsymbol{\gamma}}|_{\boldsymbol{\theta}^0} = (-p_K, \frac{1}{\beta_K^0} - p_E, -p_K, -p_E, 0)^T. \tag{10}$$

From the vectors given by (9) and (10), it becomes obvious that the FIC represents a local, rather than a global criterion, as these derivatives generally depend upon individual observations (p_K, p_E) of the regressors. This is similar to the calculation of marginal effects in non-linear models, for which one has to choose the point at which marginal effects are evaluated (see e. g. FRONDEL, VANCE, 2011).

For both the calculation of marginal effects, as well as the FIC, the sample mean $(\bar{p}_K, \bar{p}_E)$ may be a natural choice. Depending on the distribution of p_K and p_E, however, the sample mean may not be representative for the entire sample, nor does it generally equal the mean of all those values that are obtained when the FIC is evaluated at individual data points. Hence, rather than considering the mean of the explanatory variables for the calculation of the FIC, or any other average value, such as the median, one may calculate the FIC for any composition (p_K, p_E) of the regressors in the data. It can then be decided individually for any observation whether the full or the narrow model should be preferred in estimating the focus parameters. In our empirical application, we will calculate the FIC for both the sample means of the explanatory variables and for each individual observation.

3 The Concept of the FIC

As the term Focused Information Criterion suggests, it is not surprising that the FIC is based on an information matrix, which is related to FISHER's well-known information measure and represents the variance matrix of the score vector $\frac{\partial \log L}{\partial \boldsymbol{\theta}}$, where L denotes the likelihood function that is specified below for our example. This information matrix is evaluated for $\boldsymbol{\theta}_0$, that is, for the null model:

$$I^{full}|_{\boldsymbol{\theta}_0} = E[\left(\frac{\partial \log L}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}^0}\right) \cdot \left(\frac{\partial \log L}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}^0}\right)^T] = \begin{bmatrix} I_{00} & I_{01} \\ I_{10} & I_{11} \end{bmatrix}, \tag{11}$$

where the matrix entries are defined as follows:

$$
\begin{aligned}
I_{00} &:= E[(\frac{\partial \log L}{\partial \boldsymbol{\xi}}|_{\boldsymbol{\theta}^0}) \cdot (\frac{\partial \log L}{\partial \boldsymbol{\xi}}|_{\boldsymbol{\theta}^0})^T], I_{01} := E[(\frac{\partial \log L}{\partial \boldsymbol{\xi}}|_{\boldsymbol{\theta}^0}) \cdot (\frac{\partial \log L}{\partial \gamma}|_{\boldsymbol{\theta}^0})^T], \\
I_{10} &:= E[(\frac{\partial \log L}{\partial \gamma}|_{\boldsymbol{\theta}^0}) \cdot (\frac{\partial \log L}{\partial \boldsymbol{\xi}}|_{\boldsymbol{\theta}^0})^T], I_{11} := E[(\frac{\partial \log L}{\partial \gamma}|_{\boldsymbol{\theta}^0}) \cdot (\frac{\partial \log L}{\partial \gamma}|_{\boldsymbol{\theta}^0})^T].
\end{aligned}
$$

The dimension of information matrix I^{full} is (p+q) $\times$ (p+q). In our example, $p = 4$ refers to the number of parameters gathered in $\boldsymbol{\xi}$ of the null model, while $q = 5$ is the number of parameters that exclusively belong to the full model and are given by γ.

Normality assumed, the likelihood for the full model is a bivariate standard-normal density conditional on the model parameters $\boldsymbol{\theta}$ and fixed values for p_K and p_E:

$$
\begin{aligned}
L &:= L(s_L, s_K | p_K, p_E, \boldsymbol{\theta}) \\
&= \frac{1}{2\pi\sigma_K\sigma_L\sqrt{1-\rho_{KL}^2}} \exp\left[-\frac{1}{2(1-\rho_{KL}^2)}\left(\epsilon_L^2 + \epsilon_K^2 - 2\rho_{KL}\epsilon_L\epsilon_K\right)\right], \quad (12)
\end{aligned}
$$

with

$$
\begin{aligned}
\epsilon_K &:= \frac{1}{\sigma_K}(s_K - \beta_K - \beta_{KK} \cdot p_K - \beta_{KE} \cdot p_E) \sim N(0,1) \quad \text{and} \quad (13) \\
\epsilon_L &:= \frac{1}{\sigma_L}(s_L - \beta_L - \beta_{LK} \cdot p_K - \beta_{LE} \cdot p_E) \sim N(0,1). \quad (14)
\end{aligned}
$$

By taking the logarithm of (12) and differentiating with respect to the parameter vectors $\boldsymbol{\xi}$ and γ, we get the score vectors that are required for estimating the information matrix $I^{full}|_{\boldsymbol{\theta}^0}$:

$$
\begin{aligned}
\frac{\partial \log L}{\partial \boldsymbol{\xi}}|_{\boldsymbol{\theta}^0} &= \left(\frac{\partial \log L}{\partial \beta_K}, \frac{\partial \log L}{\partial \beta_L}, \frac{\partial \log L}{\partial \sigma_K}, \frac{\partial \log L}{\partial \sigma_L}\right)^T |_{\boldsymbol{\theta}^0} \\
&= \left(\frac{\varepsilon_K}{\sigma_K}, \frac{\varepsilon_L}{\sigma_L}, \frac{\varepsilon_K^2}{\sigma_K} - \frac{1}{\sigma_K}, \frac{\varepsilon_L^2}{\sigma_L} - \frac{1}{\sigma_L}\right)^T \quad (15)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial \log L}{\partial \gamma}|_{\boldsymbol{\theta}^0} &= \left(\frac{\partial \log L}{\partial \beta_{KK}}, \frac{\partial \log L}{\partial \beta_{KE}}, \frac{\partial \log L}{\partial \beta_{LK}}, \frac{\partial \log L}{\partial \beta_{LE}}, \frac{\partial \log L}{\partial \rho_{KL}}\right)^T |_{\boldsymbol{\theta}^0} \\
&= \left(\frac{\varepsilon_K p_K}{\sigma_K}, \frac{\varepsilon_K p_E}{\sigma_K}, \frac{\varepsilon_L p_K}{\sigma_L}, \frac{\varepsilon_L p_E}{\sigma_L}, \varepsilon_K \cdot \varepsilon_L\right)^T . \qquad (16)
\end{aligned}
$$

On the basis of these expressions for the score vectors, the information matrix $I^{full}|_{\boldsymbol{\theta}^0}$ is derived in detail in the appendix.

An estimate of the information matrix $I^{full}|_{\boldsymbol{\theta}^0}$ can be obtained by

$$
\hat{I}^{full}|_{\boldsymbol{\theta}^0} := \frac{1}{n}\sum_{i=1}^{n} I^{full}(p_{K,i}, p_{E,i})|_{\boldsymbol{\theta}^0}, \qquad (17)
$$

with $p_{K,i}$ and $p_{E,i}$ being the i-th observations of p_K and p_E, respectively. The estimate of $I^{full}|_{\boldsymbol{\theta}^0}$ then reads as follows:

$$
\hat{I}^{full}|_{\boldsymbol{\theta}_0} = \left[\begin{array}{cccc:ccccc}
1/\hat{\sigma}_K^2 & 0 & 0 & 0 & \bar{p}_K/\hat{\sigma}_K^2 & \bar{p}_E/\hat{\sigma}_K^2 & 0 & 0 & 0 \\
0 & 1/\hat{\sigma}_L^2 & 0 & 0 & 0 & 0 & \bar{p}_K/\hat{\sigma}_L^2 & \bar{p}_E/\hat{\sigma}_L^2 & 0 \\
0 & 0 & 2/\hat{\sigma}_K^2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2/\hat{\sigma}_L^2 & 0 & 0 & 0 & 0 & 0 \\
\hdashline
\bar{p}_K/\hat{\sigma}_K^2 & 0 & 0 & 0 & \overline{p_K^2}/\hat{\sigma}_K^2 & \frac{\overline{p_K p_E}}{\hat{\sigma}_K^2} & 0 & 0 & 0 \\
\bar{p}_E/\hat{\sigma}_K^2 & 0 & 0 & 0 & \frac{\overline{p_K p_E}}{\hat{\sigma}_K^2} & \overline{p_E^2}/\hat{\sigma}_K^2 & 0 & 0 & 0 \\
0 & \bar{p}_K/\hat{\sigma}_L^2 & 0 & 0 & 0 & 0 & \overline{p_K^2}/\hat{\sigma}_L^2 & \frac{\overline{p_K p_E}}{\hat{\sigma}_L^2} & 0 \\
0 & \bar{p}_E/\hat{\sigma}_L^2 & 0 & 0 & 0 & 0 & \frac{\overline{p_K p_E}}{\hat{\sigma}_L^2} & \overline{p_E^2}/\hat{\sigma}_L^2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right],
$$

where the unknown parameters σ_K^2 and σ_L^2 have been replaced by the corresponding ML estimates $\hat{\sigma}_K^2$ and $\hat{\sigma}_L^2$, respectively, and $\bar{p}_K$, $\bar{p}_E$, $\overline{p_K^2}$, $\overline{p_E^2}$, and $\overline{p_K p_E}$ denote the sample means of the explanatory variables, their squared values, and their cross-products, respectively.

For didactic purposes, we now present the definition of the FIC for an important special case for which the south-east block of the inverse of the information matrix is diagonal:

$$\mathbf{V} := I^{11} = (I_{11} - I_{10}I_{00}^{-1}I_{01})^{-1}, \tag{18}$$

although this does not hold true for our example. While balancing modeling bias $\mathbf{B}$ and relative estimation variability $\mathbf{V}$ (CLAESKENS, HJORT, 2003:907), for the q-dimensional case in which models may differ in q parameters $\gamma_1, ..., \gamma_q$, the FIC for a diagonal $\mathbf{V}$ is given by (see CLAESKENS, HJORT, 2003:903):

$$FIC := [\sum_{j=1}^{q} \omega_j B_j 1(\gamma_j = \gamma_j^0)]^2 + 2\sum_{j=1}^{q} \omega_j^2 V_j 1(\gamma_j \neq \gamma_j^0), \tag{19}$$

where V_j is a diagonal element of $\mathbf{V}$ and $1(\cdot)$ denotes the indicator function. The rationale underlying definition (19) is that for any deviation $\gamma_j \neq \gamma_j^0$, there is a trade-off between modeling bias, which, by definition, is zero if $\gamma_j \neq \gamma_j^0$, and relative estimation variability, which, by definition, is zero if $\gamma_j = \gamma_j^0$.

In definition (19), the purpose of inference, given by an estimate of the focus parameter μ, is taken into account in that ω_j is the jth component of a vector $\boldsymbol{\omega}$ that generally depends on focus parameter μ (see CLAESKENS, HJORT, 2003:902):

$$\boldsymbol{\omega} := I_{10}I_{00}^{-1}\frac{\partial\mu}{\partial\boldsymbol{\xi}}|_{\boldsymbol{\theta}^0} - \frac{\partial\mu}{\partial\boldsymbol{\gamma}}|_{\boldsymbol{\theta}^0}. \tag{20}$$

Finally, B_j is a component of the bias vector

$$\mathbf{B} := \sqrt{n}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^0) = \sqrt{n}\boldsymbol{\gamma}, \tag{21}$$

where the difference $\boldsymbol{\gamma} - \boldsymbol{\gamma}^0$ captures modeling bias. Recall that in our example $\boldsymbol{\gamma}^0 = \mathbf{0}$ and note that the modeling bias only vanishes if the full model were to be identical to the null model, so that $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0 = \mathbf{0}$.

For our example outlined in Section 2, the FIC simplifies for the null model to

$$FIC^0 = \omega^T \mathbf{B}\mathbf{B}^T \omega, \tag{22}$$

which represents the modeling bias of the null model relative to the full model, while the relative estimation variability vanishes for the null model by definition. In contrast, the FIC for the full model reads:

$$FIC^{full} = 2\omega^T \mathbf{V} \omega, \tag{23}$$

where $\mathbf{V}$ captures the relative estimation variability, whereas there is no modeling bias by definition: $\mathbf{B} = \mathbf{0}$.

Again referring to the special case (19), it bears noting that with a parsimonious model, the reward is a small variance contribution, $2\sum_{j=1}^{q} \omega_j^2 V_j 1(\gamma_j \neq \gamma_j^0)$, but the penalty is a larger magnitude of the term $(\sum_{j=1}^{q} \omega_j B_j 1(\gamma_j = \gamma_j^0))^2$ that originates from modeling bias. The situation is reversed for richer models. In short, including more model parameters always implies more variance, but lower bias, and vice versa.

4 Empirical Application

In this section, we apply the FIC to the well-established KLEM data set made available by Dale JORGENSON.[1] This data base has been frequently used for production analysis at the aggregate level (see e. g. JORGENSON and STIROH, 2000, FRONDEL and SCHMIDT, 2006). The data for capital, labor, energy, and materials cover 35 sectors of the U.S. economy for the years spanning 1960 to 2005. This amounts to a total of 1,610 observations. The data comprises information on real factor prices and real values of inputs to production. In addition, output prices and quantities are also included.

[1]This data set is accessible via internet: *www.economics.harvard.edu/faculty/jorgenson*.

Following our analytical example presented in Section 2, we deliberately restrict our empirical analysis to three factors, capital, labor, and energy, thereby computing the respective cost shares by subtracting the cost of materials from total cost. To keep our analysis concise, we only estimate the FIC for the null and the full model, rather than comparing the estimates of the FIC for all $2^q = 2^5 = 32$ possible (sub-)models.

Using maximum likelihood methods for estimating both the null and the full model, we calculate the FIC for our two focus measures of choice, $\eta_{K_{p_E}}$ and $\eta_{L_{p_E}}$, the cross-price elasticities of capital and labor demand both with respect to energy prices. Estimates for the FIC evaluated at the mean of the explanatory variables are displayed in Table 1. Several results bear highlighting: First, for the pooled data the FIC clearly argues in favor of the less restrictive Translog model, irrespective of the focus parameter. This does not come as a surprise, as the FIC depends on the sample size and tends to prefer the full model for abundant samples. This is also warranted from an economic perspective: When various sectors with substantially different cost structures are lumped together, the COBB-DOUGLAS model, which presumes constant cost shares (see equation system (1)), appears to be rather implausible.

Second, for both of our focus measures, the Translog model is also preferred for the majority of 22 of 35 individual industry sectors. Among these sectors are 'coal mining', 'apparel', 'lumber and wood', and 'leather'. In contrast, for eleven sectors, such as 'agriculture', 'metal mining', and 'trade', the FIC prefers COBB-DOUGLAS over Translog for both focus parameters. For these eleven sectors, therefore, any cross-price elasticity involving the price of energy can be expected to reflect the cost share of energy.

Third, for two sectors, namely 'transportation equipment' and 'electric utilities', the model choice based on the FIC depends on the focus parameter, that is, on whether the energy price elasticity of either the demand for labor or for capital is the purpose of inference.

Table 1: FIC Estimates at the Sample Means of the Explanatory Variables.

	Focus Parameter: $\eta_{L_{p_E}}$			Focus Parameter: $\eta_{K_{p_E}}$		
	FIC^0	FIC^{full}	$\mathrm{FIC}^0 > \mathrm{FIC}^{full}$	FIC^0	FIC^{full}	$\mathrm{FIC}^0 > \mathrm{FIC}^{full}$
Agriculture	0.11	1.01	0	0.11	1.36	0
Metal mining	0.12	6.80	0	0.11	10.47	0
Coal mining	16.87	4.25	1	8.31	2.59	1
Oil and gas extraction	0.37	0.22	1	3.20	0.46	1
Nonmetallic mining	3.35	1.50	1	3.02	1.11	1
Construction	9.40	1.28	1	0.22	0.04	1
Food and kindred products	10.53	5.93	1	3.10	1.65	1
Tobacco	0.52	0.23	1	1.32	0.55	1
Textile mill products	0.09	2.52	0	0.01	0.27	0
Apparel	25.42	2.25	1	1.13	0.13	1
Lumber and wood	12.45	1.61	1	2.26	0.40	1
Furniture and fixtures	1.94	2.51	0	0.13	0.14	0
Paper and allied	2.62	2.05	1	0.72	0.38	1
Printing publishing and allied	0.55	0.63	0	0.07	0.08	0
Chemicals	2.23	1.11	1	1.86	0.45	1
Petroleum and coal products	35.38	3.07	1	67.07	4.39	1
Rubber and misc. plastics	3.62	3.04	1	0.33	0.20	1
Leather	61.84	9.93	1	7.37	1.48	1
Stone clay glass	13.71	5.04	1	1.89	0.43	1
Primary metal	67.40	8.83	1	15.36	1.59	1
Fabricated metal	45.49	4.91	1	5.30	0.69	1
Machinery non-electrical	0.00	1.71	0	0.00	0.15	0
Electrical machinery	9.89	6.19	1	2.03	1.24	1
Motor vehicles	0.11	5.97	0	0.03	1.58	0
Transportation equipment	4.30	4.03	1	0.10	0.10	0
Instruments	24.21	6.02	1	0.89	0.22	1
Misc. manufacturing	3.20	7.17	0	0.55	1.33	0
Transportation	9.19	0.76	1	1.59	0.22	1
Communications	0.04	1.40	0	0.05	1.75	0
Electric utilities	0.06	0.28	0	0.24	0.08	1
Gas utilities	16.91	2.31	1	56.27	4.28	1
Trade	0.01	1.43	0	0.00	0.15	0
Finance	0.28	0.07	1	1.02	0.22	1
Services	0.00	0.90	0	0.00	0.06	0
Government enterprises	9.89	3.10	1	4.40	1.35	1
Pooled sample	400.68	9.14	1	113.31	5.43	1

Note: 1 indicates $\mathrm{FIC}^0 > \mathrm{FIC}^{full}$, that is, that the full model (Translog) is to be preferred over the null model (Cobb-Douglas). Number of observations: 1,610.

While for 'transportation equipment' the simple COBB-DOUGLAS approach appears to be appropriate for measuring the substitutability of capital and energy, the richer Translog model is preferred for measuring the substitutability of labor and energy. For the electric utilities sector, the opposite holds true. In essence, this means that no uniformly best model exists when the FIC is employed to choose among a diversity of model specifications. Rather, to appropriately capture specific features of the technology of production in these sectors, distinct model specifications have to be applied.

Table 2: Sample Means of FIC Individually Estimated at each Data Point.

	Focus Parameter: $\eta_{L_{p_E}}$			Focus Parameter: $\eta_{K_{p_E}}$		
	$\overline{\text{FIC}^0}$	$\overline{\text{FIC}^{full}}$	Share of $\text{FIC}^0 > \text{FIC}^{full}$	$\overline{\text{FIC}^0}$	$\overline{\text{FIC}^{full}}$	Share of $\text{FIC}^0 > \text{FIC}^{full}$
Agriculture	0.11	1.07	0.000	0.12	1.41	0.000
Metal mining	0.18	6.86	0.000	0.17	10.53	0.000
Coal mining	17.37	4.33	1.000	8.81	2.66	0.935
Oil and gas extraction	0.51	0.23	0.652	3.35	0.47	1.000
Nonmetallic mining	3.45	1.52	0.978	3.12	1.13	1.000
Construction	9.45	1.29	1.000	0.27	0.04	0.870
Food and kindred products	10.61	5.98	1.000	3.19	1.70	1.000
Tobacco	0.57	0.25	0.913	1.37	0.56	1.000
Textile mill products	0.10	2.53	0.000	0.01	0.28	0.000
Apparel	25.53	2.26	1.000	1.25	0.14	1.000
Lumber and wood	12.59	1.62	1.000	2.40	0.42	1.000
Furniture and fixtures	1.95	2.52	0.000	0.14	0.14	0.413
Paper and allied	2.64	2.06	0.935	0.74	0.39	0.978
Printing publishing and allied	0.56	0.64	0.130	0.07	0.08	0.239
Chemicals	2.31	1.14	1.000	1.93	0.47	1.000
Petroleum and coal products	35.52	3.08	1.000	67.21	4.40	1.000
Rubber and misc. plastics	3.64	3.04	0.935	0.35	0.21	0.913
Leather	62.22	9.99	1.000	7.74	1.55	1.000
Stone clay glass	13.80	5.06	1.000	1.99	0.44	1.000
Primary metal	67.78	8.86	1.000	15.74	1.61	1.000
Fabricated metal	45.75	4.94	1.000	5.56	0.72	1.000
Machinery non-electrical	0.00	1.72	0.000	0.00	0.15	0.000
Electrical machinery	10.00	6.25	1.000	2.15	1.30	0.891
Motor vehicles	0.11	6.02	0.000	0.04	1.63	0.000
Transportation equipment	4.31	4.04	0.935	0.10	0.11	0.370
Instruments	24.26	6.03	1.000	0.93	0.23	1.000
Misc. manufacturing	3.23	7.21	0.000	0.58	1.38	0.000
Transportation	9.40	0.77	1.000	1.79	0.23	0.935
Communications	0.04	1.42	0.000	0.05	1.78	0.000
Electric utilities	0.07	0.29	0.000	0.25	0.10	0.978
Gas utilities	17.03	2.33	1.000	56.38	4.29	1.000
Trade	0.01	1.44	0.000	0.00	0.15	0.000
Finance	0.30	0.07	1.000	1.05	0.22	1.000
Services	0.00	0.90	0.000	0.00	0.07	0.000
Government enterprizes	10.21	3.14	1.000	4.72	1.39	0.957
Pooled sample	409.04	9.40	1.000	121.66	5.69	0.995

The dependence of the model choice on the purpose of inference is more relevant if the FIC is alternatively computed for any data point individually, rather than calculating it at the sample mean of the explanatory variables. Table 2 reports the means of the FIC resulting from this exercise. There are some important discrepancies, but there is also a range of similarities to Table 1: First, for twelve sectors, the FIC unanimously prefers the Translog model for each individual observation and irrespective of the focus parameter. Among these sectors are 'apparel', 'lumber and wood', and 'leather'. Second, for nine sectors, such as 'agriculture', 'metal mining', and 'trade', the FIC always recommends COBB-DOUGLAS for any point of time and both focus parameters.

For the remaining 14 sectors, however, the application of the FIC yields mixed results. Four sectors, such as 'paper and allied' and 'transportation equipment', exhibit a mixed pattern across observations for both focus parameters, whereas for the remaining ten sectors, such as 'coal mining' or 'tobacco', the FIC uniformly prefers one model for one of either cross-price elasticities, but yields ambiguous recommendations across observations for the other focus parameter. This also applies to the pooled sample: for eight observations the COBB-DOUGLAS specification is preferred if the focus is on $\eta_{K_{p_E}}$, while Translog is always preferred for focus parameter $\eta_{L_{p_E}}$.

5 Summary and Conclusions

In choosing an appropriate model specification for describing production technologies, econometric studies on factor substitution typically resort to a few number of well-established functional forms, such as Generalized Leontief, Cobb-Douglas, or, often, Translog cost functions. In selecting a single specification out of a variety of functional forms, however, it should be recognized that one specification might be more appropriate for a certain task then another.

In this paper, we advise against selecting a production model that is preferred without any reference to the research question addressed, such as calculating a specific measure of substitutability. Rather, we suggest choosing those model specifications that fit best to the specific purposes of inference. 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. In addition to this feature, the FIC is distinguished from other model selection measures, such as AIC and SIC, 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 values of the conditioning variables. This is not a paradox, as CLAESKENS and HJORT (2003:?) note, but stems from the wish to estimate conditional expected values with optimal precision.

Using the classical example of the choice among COBB-DOUGLAS and Translog specifications and empirical data for 35 U.S. industry sectors for the time period spanning 1960 to 2005, this paper has illustrated the concept and usefulness of the FIC. When evaluated at the sample means of the explanatory variables, the model choice based on the FIC depends on the focus parameter for two out of 35 sectors. If, alternatively and more reasonably, the FIC is computed for each data point individually, the recommendations based on the FIC results are highly mixed and depend on the purpose of inference for 14 out of 35 industry sectors. This outcome, as well as the general reasoning with respect to model selection, is in line with the conclusion of FUSS, MCFADDEN and MUNDLAK (1978:241) in the context of choosing the most appropriate among a variety of substitution measures: there "is no unique natural generalization of the two factor definition ... [and] the selection of a particular definition should depend on the question asked".

Appendix

In what follows, a range of entries of information matrix $I^{full}|_{\boldsymbol{\theta}_0}(p_K, p_E)$ are calculated using definition (11), the score vectors (15) and (16), as well as $E[(\varepsilon_K)^2] = E[(\varepsilon_L)^2] = 1, E[(\varepsilon_K)^3] = E[\varepsilon_K] = E[(\varepsilon_L)^3] = E[\varepsilon_L] = 0$, while $E[\varepsilon_K\varepsilon_L] = \rho_{KL}\sigma_K\sigma_L = 0$, $E[(\varepsilon_K)^2\varepsilon_L] = E[(\varepsilon_L)^2\varepsilon_K] = 0$ for $\boldsymbol{\theta}^0$, as then $\rho_{KL} = 0$. The nine entries for the first row of $I^{full}|_{\boldsymbol{\theta}_0}$ then read as follows:

$$\begin{aligned}
E[(\frac{\partial \log L}{\partial \beta_K})^2]|_{\boldsymbol{\theta}^0} &= E[(\frac{\varepsilon_K}{\sigma_K})^2] = \frac{1}{\sigma_K^2}E[(\varepsilon_K)^2] = \frac{1}{\sigma_K^2}, \\
E[\frac{\partial \log L}{\partial \beta_K}\frac{\partial \log L}{\partial \beta_L}]|_{\boldsymbol{\theta}^0} &= E[\frac{\varepsilon_K}{\sigma_K}\frac{\varepsilon_L}{\sigma_L}] = \frac{1}{\sigma_K\sigma_L}E[\varepsilon_K\varepsilon_L] = 0, \\
E[\frac{\partial \log L}{\partial \beta_K}\frac{\partial \log L}{\partial \sigma_K}]|_{\boldsymbol{\theta}^0} &= E[\frac{\varepsilon_K}{\sigma_K}(\frac{\varepsilon_K^2}{\sigma_K} - \frac{1}{\sigma_K})] = \frac{1}{\sigma_K^2}(E[(\varepsilon_K)^3] - E[\varepsilon_K]) = 0, \\
E[\frac{\partial \log L}{\partial \beta_K}\frac{\partial \log L}{\partial \sigma_L}]|_{\boldsymbol{\theta}^0} &= E[\frac{\varepsilon_K}{\sigma_K}(\frac{\varepsilon_L^2}{\sigma_L} - \frac{1}{\sigma_L})] = \frac{1}{\sigma_K\sigma_L}(E[\varepsilon_K(\varepsilon_L)^2] - E[\varepsilon_K]) = 0, \\
E[\frac{\partial \log L}{\partial \beta_K}\frac{\partial \log L}{\partial \beta_{KK}}]|_{\boldsymbol{\theta}^0} &= E[\frac{\varepsilon_K}{\sigma_K}\frac{\varepsilon_K p_K}{\sigma_K}] = \frac{p_K}{\sigma_K^2}E[(\varepsilon_K)^2] = \frac{p_K}{\sigma_K^2}, \\
E[\frac{\partial \log L}{\partial \beta_K}\frac{\partial \log L}{\partial \beta_{KE}}]|_{\boldsymbol{\theta}^0} &= E[\frac{\varepsilon_K}{\sigma_K}\frac{\varepsilon_K p_E}{\sigma_K}] = \frac{p_E}{\sigma_K^2}E[(\varepsilon_K)^2] = \frac{p_E}{\sigma_K^2}, \\
E[\frac{\partial \log L}{\partial \beta_K}\frac{\partial \log L}{\partial \beta_{LK}}]|_{\boldsymbol{\theta}^0} &= E[\frac{\varepsilon_K}{\sigma_K}\frac{\varepsilon_L p_K}{\sigma_L}] = \frac{p_K}{\sigma_K\sigma_L}E[\varepsilon_K\varepsilon_L] = 0, \\
E[\frac{\partial \log L}{\partial \beta_K}\frac{\partial \log L}{\partial \beta_{LE}}]|_{\boldsymbol{\theta}^0} &= E[\frac{\varepsilon_K}{\sigma_K}\frac{\varepsilon_L p_E}{\sigma_L}] = \frac{p_E}{\sigma_K\sigma_L}E[\varepsilon_K\varepsilon_L] = 0, \\
E[\frac{\partial \log L}{\partial \beta_K}\frac{\partial \log L}{\partial \rho_{KL}}]|_{\boldsymbol{\theta}^0} &= E[\frac{\varepsilon_K}{\sigma_K}\varepsilon_K\varepsilon_L] = \frac{1}{\sigma_K}E[(\varepsilon_K)^2\varepsilon_L] = 0.
\end{aligned}$$

While, by analogy, the second row of $I^{full}|_{\boldsymbol{\theta}_0}$ is similar to the first, and is thus not calculated explicitly here, the unknown entries of the third row, beginning from the diagonal element for symmetry reasons, are given by:

$$\begin{aligned}
E[(\frac{\partial \log L}{\partial \sigma_K})^2]|_{\boldsymbol{\theta}^0} &= E[(\frac{\varepsilon_K^2}{\sigma_K} - \frac{1}{\sigma_K})^2] = E[\frac{(\varepsilon_K)^4}{\sigma_K^2} - 2\frac{(\varepsilon_K)^2}{\sigma_K^2} + \frac{1}{\sigma_K^2}] \\
&= \frac{1}{\sigma_K^2}E[(\varepsilon_K)^4 - 2(\varepsilon_K)^2 + 1] = \frac{1}{\sigma_K^2}[3 - 2 + 1] = \frac{2}{\sigma_K^2},
\end{aligned}$$

$$
\begin{aligned}
E[\frac{\partial \log L}{\partial \sigma_K}\frac{\partial \log L}{\partial \sigma_L}]|_{\boldsymbol{\theta}^0} &= E[(\frac{\varepsilon_K^2}{\sigma_K}-\frac{1}{\sigma_K})(\frac{\varepsilon_L^2}{\sigma_L}-\frac{1}{\sigma_L})] = \frac{1}{\sigma_K\sigma_L}E[(\varepsilon_K\varepsilon_L)^2-(\varepsilon_K)^2-(\varepsilon_L)^2+1] \\
&= \frac{1}{\sigma_K\sigma_L}[1-1-1+1]=0, \\
E[\frac{\partial \log L}{\partial \sigma_K}\frac{\partial \log L}{\partial \beta_{KK}}]|_{\boldsymbol{\theta}^0} &= E[(\frac{\varepsilon_K^2}{\sigma_K}-\frac{1}{\sigma_K})\frac{\varepsilon_K p_K}{\sigma_K}] = \frac{p_K}{\sigma_K^2}E[(\varepsilon_K)^3-\varepsilon_K]=0, \\
E[\frac{\partial \log L}{\partial \sigma_K}\frac{\partial \log L}{\partial \beta_{KE}}]|_{\boldsymbol{\theta}^0} &= E[(\frac{\varepsilon_K^2}{\sigma_K}-\frac{1}{\sigma_K})\frac{\varepsilon_K p_E}{\sigma_K}] = \frac{p_E}{\sigma_K^2}E[(\varepsilon_K)^3-\varepsilon_K]=0, \\
E[\frac{\partial \log L}{\partial \sigma_K}\frac{\partial \log L}{\partial \beta_{LK}}]|_{\boldsymbol{\theta}^0} &= E[(\frac{\varepsilon_K^2}{\sigma_K}-\frac{1}{\sigma_K})\frac{\varepsilon_L p_K}{\sigma_L}] = \frac{p_K}{\sigma_K\sigma_L}E[(\varepsilon_K)^2\varepsilon_L-\varepsilon_L]=0, \\
E[\frac{\partial \log L}{\partial \sigma_K}\frac{\partial \log L}{\partial \beta_{LE}}]|_{\boldsymbol{\theta}^0} &= E[(\frac{\varepsilon_K^2}{\sigma_K}-\frac{1}{\sigma_K})\frac{\varepsilon_L p_E}{\sigma_L}] = \frac{p_E}{\sigma_K\sigma_L}E[(\varepsilon_K)^2\varepsilon_L-\varepsilon_L]=0, \\
E[\frac{\partial \log L}{\partial \sigma_K}\frac{\partial \log L}{\partial \rho_{KL}}]|_{\boldsymbol{\theta}^0} &= E[(\frac{\varepsilon_K^2}{\sigma_K}-\frac{1}{\sigma_K})\varepsilon_K\varepsilon_L] = \frac{1}{\sigma_K}E[(\varepsilon_K)^3\varepsilon_L-\varepsilon_K\varepsilon_L]=0,
\end{aligned}
$$

as for $\boldsymbol{\theta}^0$ and thus $\rho_{KL}=0$, $E[(\varepsilon_K\varepsilon_L)^2]=E[(\varepsilon_K)^2]\cdot E[(\varepsilon_L)^2]=1$ and $E[(\varepsilon_K)^3\varepsilon_L]=E[(\varepsilon_K)^3]\cdot E[\varepsilon_L]=0$. Again, by analogy, the fourth row resembles the third row, so that its explicit calculation can be omitted. Next, the five unknown entries of the fifth row read as follows:

$$
\begin{aligned}
E[(\frac{\partial \log L}{\partial \beta_{KK}})^2]|_{\boldsymbol{\theta}^0} &= E[(\frac{\varepsilon_K p_K}{\sigma_K})^2] = \frac{p_K^2}{\sigma_K^2}E[(\varepsilon_K)^2] = \frac{p_K^2}{\sigma_K^2}, \\
E[\frac{\partial \log L}{\partial \beta_{KK}}\frac{\partial \log L}{\partial \beta_{KL}}]|_{\boldsymbol{\theta}^0} &= E[(\frac{\varepsilon_K p_K}{\sigma_K})(\frac{\varepsilon_K p_E}{\sigma_K})] = \frac{p_K p_E}{\sigma_K^2}E[(\varepsilon_K)^2] = \frac{p_K p_E}{\sigma_K^2}, \\
E[\frac{\partial \log L}{\partial \beta_{KK}}\frac{\partial \log L}{\partial \beta_{LK}}]|_{\boldsymbol{\theta}^0} &= E[(\frac{\varepsilon_K p_K}{\sigma_K})(\frac{\varepsilon_L p_K}{\sigma_L})] = \frac{(p_K)^2}{\sigma_K\sigma_L}E[\varepsilon_K\varepsilon_L]=0, \\
E[\frac{\partial \log L}{\partial \beta_{KK}}\frac{\partial \log L}{\partial \beta_{LE}}]|_{\boldsymbol{\theta}^0} &= E[(\frac{\varepsilon_K p_K}{\sigma_K})(\frac{\varepsilon_L p_E}{\sigma_L})] = \frac{p_K p_E}{\sigma_K\sigma_L}E[\varepsilon_K\varepsilon_L]=0, \\
E[\frac{\partial \log L}{\partial \beta_{KK}}\frac{\partial \log L}{\partial \rho_{KL}}]|_{\boldsymbol{\theta}^0} &= E[(\frac{\varepsilon_K p_K}{\sigma_K})(\varepsilon_K\varepsilon_L)] = \frac{p_K}{\sigma_K}E[(\varepsilon_K)^2\varepsilon_L]=0.
\end{aligned}
$$

While the remaining rows can be calculated accordingly, the last entry of the last row is given by:

$$
E[(\frac{\partial \log L}{\partial \rho_{KL}})^2]|_{\boldsymbol{\theta}^0} = E[(\varepsilon_K\varepsilon_L)^2] = E[(\varepsilon_K)^2]\cdot E[(\varepsilon_L)^2] = 1.
$$

References

AKAIKE, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6): 716-723.

APOSTOLAKIS, B. E. (1990) Energy-Capital-Substitutability/Complementarity. *Energy Economics*, **12 (1)**: 48-58.

BEHL, P. , DETTE, H. , FRONDEL M. , TAUCHMANN, H. (2010) Choice is Suffering: A Focused Information Criterion for Model Selection. Discussion Paper Series of SFB 823, No. 42/10, Dortmund.

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.

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*, 88 (1), 186-192.

FRONDEL, M., SCHMIDT, C. M. (2004) Facing the Truth about Separability: Nothing Works Without Energy. *Ecological Economics*, 51 (3-4), 217-223.

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

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

FRONDEL, M. , VANCE, C. (2011) Interpreting the Outcomes of Two-Part Models. *Applied Economics Letters*, forthcoming.

FUSS, M., MCFADDEN, D., MUNDLAK, Y. (1978) A Survey of Functional Forms in the Economic Analyses of Production. In: *Production Economics: A Dual Approach to Theory and Applications.* FUSS M, MCFADDEN D. (editors). North-Holland, **1**: 219-68, Amsterdam.

JORGENSON, D. W. , STIROH, K. J. (2000) Raising the Speed Limit: U.S. Economic Growth in the Information Age. *Brookings Papers on Economic Activity* 1, 125-211.

KINTIS, A. , PANAS, E. E. (1989) The Capital-Energy Controversy: Further results. *Energy Economics*, **11**: 201-212.

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, 461-464.