# Model selection *

Jean-Marie Dufour †
Université de Montréal

June 27, 2007, 12:22pm

Forthcoming in the *New Palgrave Dictionary of Economics,* $2^n d$ *Edition*.

* This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec), and a Killam Fellowship (Canada Council for the Arts).

† Canada Research Chair Holder (Econometrics). Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL: 1 (514) 343 2400; FAX: 1 (514) 343 5831; e-mail: jean.marie.dufour@umontreal.ca . Web page: http://www.fas.umontreal.ca/SCECO/Dufour

## ABSTRACT

The problem of statistical model selection in econometrics and statistics is reviewed. Model selection is interpreted as a decision problem through which a statistical model is selected in order to perform statistical analysis, such as estimation, testing, confidence set construction, forecasting, simulation, policy analysis, etc. Broad approaches to model selection are described: (1) hypothesis testing procedures, including specification and diagnostic tests; (2) penalized goodness-of-fit methods, such as information criteria; (3) Bayesian approaches; (4) forecast evaluation methods. The effect of model selection on subsequent statistical inference is also discussed.



**Key words:** model selection, specification error, hypothesis test, specification test, information criterion, Akaike, Bayesian statistics, forecast evaluation, endogeneity and exogeneity, forecasting, serial correlation, statistical inference, structural change, time series analysis.

**Journal of Economic Literature classification**: C1, C12, C5, C51, C44.

# Contents

The purpose of econometric analysis is to develop mathematical representations of observable phenomena, which we call *models* or *hypotheses* (models subject to restrictions). Such models are then used to perform parameter estimation, test hypotheses, build confidence sets, make forecasts, conduct simulations, analyze policies, etc. A central feature of modelling activity is the fact that models are usually interpreted as stylized (or simplified) representations that can perform certain tasks – such as prediction – but (eventually) not others, and they are treated *as if they were true* for certain purposes. Indeed, summarizing and stylizing observed phenomena can be viewed as essential components of modelling activity, which make it useful. This feature is not specific to economics and is shared by other sciences [see Cartwright (1983)].

Models can be classified as either *deterministic* or *stochastic*. *Deterministic models*, which often claim to make arbitrarily precise predictions, can be useful in theoretical activity. However, such models are rarely viewed as appropriate representations of observed data; for example, unless they are highly complex or indeterminate, they are typically logically inconsistent with data. For this reason, models used for econometric analysis are usually *stochastic* (or *statistical*).

Formally, a statistical model is a family of probability distributions (or measures) which are proposed to represent observed data. *Model selection*, in this context, is the task of selecting a family of proposed probability distributions, which will then be used to analyze data and perform other statistical inference operations (such as parameter estimation, hypothesis testing, etc.).

A basic feature of probability models is that they are typically *unverifiable*: as for any theory that makes an indefinite number of predictions, we can never be sure that the model will not be at odds with new data. Moreover, they are *logically unfalsifiable*: in contrast with deterministic models, a probabilistic model is usually logically compatible with all possible observation sets. Consequently, model selection can depend on a wide array of elements, such as the objectives of the model, (economic) theory, the data themselves, and various conventions.

Features which are often viewed as desirable include: (1) simplicity or *parsimony* [Zellner, Keuzenkamp and McAleer (2001)]; (2) the ability to deduce testable (or falsifiable) hypotheses [Popper (1968)]; (3) the possibility of interpreting model parameters in terms of economic theory, if not consistency with economic theory; (4) the ability to satisfactorily perform the tasks for which the model is built (prediction, for example); (5) consistency with observed data. It is important to note these characteristics depend (at least, partially) on conventional elements, such as the objectives of the model, criteria upon which a model will be deemed "satisfactory", etc. For further discussions of these general issues, the reader may consult Poirier (1994), Morgan and Morrison (1999), Keuzenkamp (2000), Zellner et al. (2001) and Dufour (2003).

In this article, we focus on statistical methods for selecting a model on the basis of the available data. Methods for that purpose can be classified in four broad (not mutually exclusive) categories;

1. hypothesis testing procedures, including specification and diagnostic tests;
2. penalized goodness-of-fit methods, such as information criteria;
3. Bayesian approaches;
4. forecast evaluation methods.

The three first approaches are meant to be applicable "in-sample", while the last approach *stricto sensu* requires observations that are not available when the model is selected, but may lead to model revision.

For general reviews of the topic of statistical model selection in econometrics and statistics, see Hocking (1976), Leamer (1978, 1983), Draper and Smith (1981), Judge, Griffiths, Carter Hill, Lütkepohl and Lee (1985, Chapters 7 and 21), Sakamoto, Ishiguro and Kitagawa (1985), Grasa (1989), Choi (1992), Gouriéroux and Monfort (1995, Chapter 22), Charemza and Deadman (1997), McQuarrie and Tsai (1998), Burnham and Anderson (2002), Clements and Hendry (2002), Miller (2002), and Bhatti, Al-Shanfari and Hossain (2006). It is also interesting to note that classification techniques in statistics contain results that may be relevant to model selection. This topic, however, goes beyond the scope of the present article; for further discussion, see Krishnaiah and Kanal (1982).

# 1. Model selection and specification errors

Most model selection methods deal in different ways with a trade-off between model *realism* – which usually suggests considering relatively general, hence complex models – and *parsimony*. From the viewpoint of estimation, for example, a model which is too simple (or parsimonious) involves *specification errors* and *biases* in parameter estimation, while too complex a model leads to parameter estimates with large variances. If the objective is forecasting, it is usually unclear which effect dominates.

For example, let us consider a linear regression model of the form:

$$y_t = x_{t1}\beta_1 + x_{t2}\beta_2 + \cdots + x_{tk}\beta_k + u_t\,, \quad t = 1,\, \dots\,,\, T, \tag{1.1}$$

where $y_t$ is a dependent variable and $x_{t1},\, \dots\,,\, x_{tk}$ are explanatory variables, $u_t$ is a random disturbance which is typically assumed to be independent of (or uncorrelated with) the explanatory variables. In the classical linear model, it is assumed that the regressors can be taken as fixed and that the disturbances $u_1,\, \dots\,,\, u_T$ are independent and identically distributed (i.i.d) according to a $N(0, \sigma^2)$ distribution. In this context, model selection typically involves selecting the regressors to be included as well as various distributional assumptions to be made upon the disturbances.

An especially important version of (1.1) is the autoregressive model:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + u_t\,, \quad t = 1, \ldots, T. \tag{1.2}$$

Then a central model selection issue consists in setting the order $p$ of the process. In such models, there is typically little theoretical guidance on the order, so data-based order selection rules can be quite useful. A related setup where model selection is usually based on statistical methods is the class of autoregressive-moving-average models (ARMA):

$$y_t = \beta_0 + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + u_t - \theta_1 u_{t-1} + \cdots + \theta_1 u_{t-q}\,, \tag{1.3}$$

where the orders $p$ and $q$ must be specified.

By considering the simple linear regression model, it is easy to see that excluding irrelevant variables can lead to biases in parameter estimates [Theil (1957)]. On the other hand, including irrelevant regressors raises the variances of the estimators. The overall effect on the mean square error (MSE) of the estimator and, more generally, how closely it will tend to approach the parameter value may be ambiguous. It is well known that a biased estimator may have lower MSE than an unbiased estimator. This may be particularly important in forecasting where a simple "false" model may easily provide better forecasts than a complicated "true" model, because the latter may be affected by imprecise parameter estimates [Allen (1971)].

# 2. Hypothesis testing approaches

Since hypothesis tests are based on a wide body of statistical theory [see Lehmann (1986), Gouriéroux and Monfort (1995)], such procedures are widely used for assessing, comparing and selecting models. Furthermore, econometric models are also based on economic theory which suggests basic elements that can be used for specifying models. This entails a form of asymmetry, in which restrictions suggested by economic theory will be abandoned only if "sufficient evidence" becomes available. Although significance tests are meant to decide whether a given hypothesis (which usually takes the form of a restricted model) is compatible with the data, such procedures can also be used for model selection. It is interesting to note that the methodology originally proposed by Box and Jenkins (1976) for specifying ARMA models was almost exclusively based on significance tests (essentially, autocorrelation tests).

There are two basic ways of using hypothesis tests for that purpose. The first one is *forward* or *specific-to-general* approach, in which one starts from a relatively simple model and then checks whether the model can be deemed "satisfactory". This typically involves various specification tests, such as:

1. residual-based tests, including tests for heteroskedasticity, autocorrelation, outliers, distributional assumptions (e.g., normality), etc.;
2. tests for unit roots and/or stationarity, to decide whether corrections for integrated variables may be needed;
3. tests for the presence of structural change;
4. exogeneity tests, to decide whether corrections for endogeneity – such as instrumental variable (IV) methods – are required;
5. tests for the addition of explanatory variables;
6. tests of the functional form used (e.g., linearity vs. nonlinearity).

There is a considerable literature on specification tests in econometrics; see Godfrey (1988), MacKinnon (1992) and Davidson and MacKinnon (1993). Systematic procedures for adding variables are also know in statistics as *forward selection* or *stepwise regression* procedures [Draper and Smith (1981)].

The second way is the *backward* or *general-to-specific* approach, in which one starts from a relatively comprehensive model which includes all the relevant variables. This model is then simplified by checking which variables are significant. *Backward selection* procedures in statistics [Draper and Smith (1981)] and the general-to-specific approach in econometrics [Davidson, Hendry, Srba and Yeo (1978), Charemza and Deadman (1997)] can be viewed as illustrations of this approach.

In practical work, the backward and forward approaches are typically combined. Both involve a search for a model which is both parsimonious and consistent with the data. However, the results may differ. Specifying a model through significance tests involves many judgments and depends on idiosyncratic decisions. Further, standard hypothesis tests involve the use of typically conventional levels (such as the commonly used $5\%$ level). The powers of the tests can also have a strong influence on the results.

# 3. Penalized goodness-of-fit criteria

As pointed out by Akaike (1974), it is not clear that hypothesis testing is a good basis for model selection. Instead, the problem of model selection may be better interpreted as an estimation problem involving a well-defined loss function. This leads to the topic of goodness-of-fit criteria.

A common way of assessing the performance of a regression model, such as (1.1), consists in computing the coefficient of determination, *i.e.* the proportion of the dependent variable variance which is "explained" by the model:

$$R^2 = 1 - \frac{\hat{V}(u)}{\hat{V}(y)} \tag{3.1}$$

where $\hat{V}(u) = \sum_{t=1}^{T} \hat{u}_t^2/T$, $\hat{V}(y) = \sum_{t=1}^{T} (y_t - \overline{y})^2/T$, $\overline{y} = \sum_{t=1}^{T} y_t/T$ and $\hat{u}_1, \ldots, \hat{u}_T$ are least squares residuals. This measure, however, has the inconvenient feature that it always increases when a variable is added to the model, even if it is completely irrelevant, and it can be made equal to its maximal value of one by including a sufficient number of regressors (for example, using any set of $T$ linearly independent regressors).

An early way of avoiding this problem was proposed by Theil (1961, p. 213) who suggested that $\hat{V}(u)$ and $\hat{V}(y)$ be replaced by the corresponding unbiased estimators $s^2 = \sum_{t=1}^{T} \hat{u}_t^2/(T-k)$ and $s_y^2 = \sum_{t=1}^{T} (y_t - \overline{y})^2/(T-1)$. This yields the adjusted coefficient of determination:

$$\overline{R}^2 = 1 - \frac{s^2}{s_y^2} = 1 - \frac{T-1}{T-k}\left(1 - R^2\right) = R^2 - \frac{k-1}{T-k}\left(1 - R^2\right).$$

It is easy to see that $\overline{R}^2$ may increase when the number of regressors increases. Note that maximizing $\overline{R}^2$ is equivalent to minimizing the "unbiased estimator" $s^2$ of the disturbance variance. Further, if two regression models (which satisfy the assumptions of the classical linear model) are compared, and if one of these is the "true" model, then the value of $s^2$ associated with the true model is smaller on average than the one of the other model [see Theil (1961, p. 543)]. On the other hand, in large samples, the rule which consists in maximizing $\overline{R}^2$ does not select the true model with a probability converging to one, *i.e.* it is not consistent [see Gouriéroux and Monfort (1995, section 2.3)].

Another approach consists in evaluating the "distance" between the selected model and the true (unknown) model. Let $f(y)$ the density associated with the postulated model and $f_o(y)$ the density of the true model, where $Y = (y_1, \ldots, y_T)'$. One such distance is the *Kullback distance*:

$$\begin{aligned} I(f, f_o) &= \int \log\left[f_o(y)/f(y)\right] f_o(y)\, dy \\ &= \underset{f_o}{E}\left\{\log\left[f_o(Y)/f(Y)\right]\right\} \\ &= \underset{f_o}{E}\left\{\log\left[f_o(Y)\right]\right\} - \underset{f_o}{E}\left\{\log\left[f(Y)\right]\right\}. \end{aligned}$$

Minimizing $I(f, f_o)$ with respect to $f$ is equivalent to minimizing $-\underset{f_o}{E}\left\{\log\left[f(Y)\right]\right\}$. We obtain an information criterion by selecting an "estimator" of $\underset{f_o}{E}\{\log[f(Y)]\}$.

For the case where the model is estimated by maximizing a likelihood function $L_T(\theta)$ over a $K \times 1$ parameter vector $\theta$, Akaike (1973) suggests that $L(\hat{\theta})$ can be viewed as a natural estimator of $\underset{f_o}{E}\{\log[f(Y)]\}$. However, the fact that $\theta$ has been estimated introduces a bias. This bias is (partially) corrected – using an expansion argument – by subtracting the number $K$ from $L(\hat{\theta})$. This suggests the following information criterion:

$$AIC_L(\hat{\theta}_T) = -2L_T(\hat{\theta}_T) + 2K \tag{3.2}$$

where $K$ is the dimension of $\theta$ (the number of estimated parameters) and multiplication by 2 is introduced to simplify the algebra. Among a given set of models, the one with the lowest $AIC$ is selected.

The above criterion has also been generalized by various authors leading the following general class of criteria:

$$IC_L(\hat{\theta}_T) = -2L_T(\hat{\theta}_T) + c(T,\, K)K \tag{3.3}$$

where $c(T,\, K)$ is a function of $T$ and $K$. In the case of Gaussian likelihoods, such as (1.1) or (1.2) with i.i.d. $N(0, \sigma^2)$ disturbances, we have $L_T(\hat{\theta}_T) = -(T/2)\ln(\hat{\sigma}_T^2) + d_T$, where $d_T$ is a constant which only depends on $T$, so that minimizing $IC_L(\hat{\theta}_T)$ is equivalent to minimizing

$$IC(\hat{\theta}_T) = \ln(\hat{\sigma}_T^2) + c(T,\, K)\frac{K}{T}\,. \tag{3.4}$$

Alternative values of $c(T,\, K)$ which have been proposed include:

1. $c(T,\, K) = 2$ [Akaike (1969)], which yields what is usually called the AIC criterion;
2. $c(T,\, K) = \ln(T)$ [Schwarz (1978)];
3. $c(T,\, K) = 2\delta_T \ln(\ln T)$ where $\limsup_{T\to\infty} \delta_T > 1$ [Hannan and Quinn (1979)];
4. $c(T,\, K) = 2 + \frac{2K(K+1)}{T-K-1}$ [Hurvitch and Tsai (1989)], which leads to the AIC$_c$ criterion.

An especially convenient feature of such information criteria is the fact that they can be applied to both regression models [through (3.4)] as well as to various nonlinear models [using (3.3)].

Other related rules include: (1) criteria based on an estimate of the *final prediction error*, which try to estimate the mean square prediction error taking into account estimation uncertainty [Akaike (1969, 1970), Mallows (1973), Amemiya (1980)]; (2) the *criterion* *autoregressive transfer* (CAT) function proposed by Parzen (1977) for selecting the order of an autoregressive process; (3) Sawa's (1978) Bayesian information criterion (BIC).

By far, the information criteria are the most widely used in practice. Some theoretical (non-)optimality properties have been established. In particular, when one of the models compared is the "true" one, it was observed by Shibata (1976) that Akaike's criterion is not consistent, in the sense that it does not select the most parsimonious true model with probability converging to one (as the sample size goes to infinity). Instead, even in large samples, it has a high probability of picking a model with "too many parameters". By contrast, the criterion proposed by Hannan and Quinn (1979) is consistent under fairly general conditions, which also entails that Schwarz's (1978) criterion also leads to consistent model selection. On the other hand, the AIC criterion has a different optimality property, in the sense that it tends minimize the one-step expected quadratic forecast error [Shibata (1980)].

On consistency, it is also interesting to observe that consistent model selection rules can be obtained provided each model is tested through a consistent test procedure (against all the other models considered) and the level of the test declines with the sample size at an appropriate rate (which depends on the asymptotic behavior of the test statistic); see Pötscher (1983).

Model selection criteria of the information type have the advantage of being fairly mechanical. On the other hand, they can be become quite costly to apply in practice when the number of models considered is large.

# 4. Bayesian model selection

Bayesian model selection involves comparing models through their "posterior probabilities" giving observed data. Suppose we have two models $M_1$ and $M_2$ each of which postulates that the observation vector $\mathbf{y}$ follows a probability density which depends on a parameter vector: $p_{\mathbf{y}}(y|\theta_1, M_1)$ under $M_1$, and $p_{\mathbf{y}}(y|\theta_2, M_2)$ under $M_2$, where $\theta_1$ and $\theta_2$ are unknown parameter vectors (which may have different dimensions). Further, each one of the parameter vector is assigned a "prior distribution" $[p(\theta_1|M_1)$ and $p(\theta_2|M_2)]$, and each model a "prior probability" $[p(M_1)$ and $p(M_2)]$. Then one may compute the "posterior probability" of each model given the data:

$$p(M_i|\mathbf{y}) = p(M_i) \int p_{\mathbf{y}}(\mathbf{y}|\theta_i,\, M_1) p(\theta_i|M_i)\, d\theta_i\ , \quad i = 1,\, 2. \tag{4.1}$$

This posterior probability of each model provides a direct measure of the "plausibility" of each model. In such contexts, the ratio

$$K_{12} = \frac{p(M_1|\mathbf{y})}{p(M_2|\mathbf{y})} \tag{4.2}$$

is called the "posterior odds ratio" of $M_1$ relative to $M_2$.

A rational decision rule for selecting between $M_1$ and $M_2$ then emerges if we can specify a loss function such as:

$$L(i,\, j) = \text{ cost of choosing } M_j \text{ when } M_i \text{ is true.} \tag{4.3}$$

If $L(i, i) = 0$ for $i = 1, 2$, expected loss is minimized by choosing $M_1$ when

$$K_{12} \geq \frac{L(2,\, 1)}{L(1,\, 2)}\,, \tag{4.4}$$

and $M_2$ when otherwise. In particular, if $L(1,\, 2) = L(2,\, 1)$, expected loss is minimized by choosing the model with the highest posterior probability. Such rules can be extended to problems where more than two models are compared.

The Bayesian approach automatically introduces a penalty for non-parsimony and easily allows the use of decision-theoretic considerations. The main difficulty consists in assigning prior distributions on model parameters and prior probabilities to competing models. For further discussion, see Zellner (1971, Chapter X), Leamer (1978, 1983), Gelman, Carlin, Stern and Rubin (2003) and Lancaster (2004).

# 5. Forecast evaluation

In view of the fact that forecasting is one of the most common objectives for building econometric models, alternative models are often assessed by studying *post-sample* forecasts. Three types of assessments are typically considered in such contexts: (1) tests of predictive failure; (2) descriptive measures of forecast performance, which can be compared across models; (3) tests of predictive ability.

A test of predictive failure involves testing whether the prediction errors associated with a model are consistent with the model. This suggests testing whether forecasts are "unbiased" or "too large" to be consistent with the model. The well-known predictive test for structural change proposed by Chow (1960) is an early example of such an approach. For further discussion and extensions, see Box and Tiao (1976), Dufour (1980), Pesaran, Smith and Yeo (1985), Dufour, Ghysels and Hall (1994), Dufour and Ghysels (1996) and Clements and Hendry (1998).

Common measures of forecast performance involve mean errors, mean square errors, mean absolute errors, etc.; see Theil (1961) and Diebold (2004). Although commonly used, such measures are mainly descriptive. They can usefully be complemented by tests of predictive ability. Such procedures test whether the difference between expected measures of forecast performance is zero (or less than zero) against an alternative where it is different from zero (or larger than zero). Tests of this type were proposed, among others, by Meese and Rogoff (1988), Diebold and Mariano (1995), Harvey, Leybourne and Newbold (1997), West (1996), West and McCracken (1998) and White (2000); for reviews, see also Mariano (2002) and McCracken and West (2002).

It is important to note that predictive performance and predictive accuracy depend on two features: first, whether the theoretical model used is close to the unknown data distribution and second, the ability to estimate accurately model parameters (hence on sample size available for estimating these). For a given sample size, a false but parsimonious model may well have better predictive ability than the "true" model.

# 6. Post-model selection inference

An important issue often raised in relation with model selection is its effect on the validity of inference – such as estimation, tests and confidence sets – obtained after a process of model selection (or pre-testing). This issue is subtle and complex. Not surprisingly, both positive and negative assessments can be found.

On the positive side, it has been observed that pre-testing (or model selection) does allow one to produce so-called "superefficient" (or Hodges) estimators, whose asymptotic variance can be at least as low as the Cramér-Rao efficiency bound and lower at certain points; see Le Cam (1953). This may be viewed as a motivation for using consistent pre-testing.

Furthermore, consistent model selection does not affect the asymptotic distributions of various estimators and test statistics, so the asymptotic validity of inferences based on a model selected according to such a rule is maintained; see Pötscher (1991) and Dufour et al. (1994).

On the negative side, it is important to note that these are only asymptotic results. In particular, these are pointwise convergence results, not uniform convergence results, so they may be quite misleading concerning what happens in finite samples; for some examples, see Dufour (1997) and Pötscher (2002). For estimation, there is a considerable literature on the finite-sample distribution of pre-test estimators, which can be quite different of their limit distributions [Judge and Bock (1978), Danilov and Magnus (2004)]. For a critical discussion of the effect of model selection on tests and confidence sets, see Leeb and Pötscher (2005).

# 7. Conclusion

The problem of model selection is one of the most basic and challenging problems of statistical analysis in econometrics. Much progress has been done in recent years in developing better model selection procedures and for understanding the consequences of model selection.

But model building remains largely an art in which subjective judgements play a central role. Developing procedures applicable to complex models, which may involve a large number of candidate variables, and allowing for valid statistical inference in the presence of model selection remain difficult issues to which much further research should be devoted.

**See also**: Bayesian statistics; econometrics; endogeneity and exogeneity; forecasting; heteroskedasticity and autocorrelation corrections; linear models; models; serial correlation; specification problems in econometrics; statistical decision theory; statistical inference; structural change; testing; time series analysis.

**Index terms**: ARMA models; autocorrelation; Bayesian statistics; deterministic models; econometrics; endogeneity; forecasting; forecast evaluation; heteroskedasticity; linear models; model selection; models; parsimony; probability models; serial correlation; specification problems in econometrics; statistical decision theory; statistical inference; stochastic models; structural change; testing; time series analysis.

# References


Akaike, H. (1969), 'Fitting autoregressive models for prediction', *Annals of the Institute of Statistical Mathematics* **21**, 243–247.

Akaike, H. (1970), 'Statistical predictor identification', *Annals of the Institute of Statistical Mathematics* **22**, 203–217.

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

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **AC-19**, 716–723.

Allen, D. M. (1971), 'Mean square error prediction as a criterion for selecting variables', *Technometrics* **13**, 469–475.

Amemiya, T. (1980), 'Selection of regressors', *International Economic Review* **21**, 331–354.

Bhatti, M. I., Al-Shanfari, H. and Hossain, M. Z. (2006), *Econometric Analysis of Model Selection and Model Testing*, Ashgate, Aldershot, U.K.

Box, G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control*, second edn, Holden-Day, San Francisco.

Box, G. E. P. and Tiao, G. C. (1976), 'Comparison of forecast and actuality', *Applied Statistics* **64**, 195–200.

Burnham, K. P. and Anderson, D. R. (2002), *Model Selection and Multi-Model Inference: A Practical Information Theoretic Approach*, Springer-Verlag, New York.

Cartwright, N. (1983), *How the Laws of Physics Lie*, Oxford University Press, Oxford, U.K.

Charemza, W. W. and Deadman, D. F. (1997), *New Directions in Econometric Practice: General to Specific Modelling, Cointegration and Vector Autoregression*, second edn, Edward Elgar, Aldershot, U.K.

Choi, B. (1992), *ARMA Model Identification*, Springer-Verlag, New York.

Chow, G. C. (1960), 'Tests of equality between sets of coefficients in two linear regressions', *Econometrica* **28**, 591–605.

Clements, M. P. and Hendry, D. F. (1998), *Forecasting Economic Time Series*, Cambridge University Press, Cambridge, U.K.

Clements, M. P. and Hendry, D. F., eds (2002), *A Companion to Economic Forecasting*, Blackwell Companions to Contemporary Economics, Blackwell, Oxford, U.K.

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

Davidson, J. E. H., Hendry, D. F., Srba, F. and Yeo, S. (1978), 'Econometric modelling of the aggregate time-serie relationship between consumers' expenditure and income in the United Kingdom', *Economic Journal* **88**, 661–692.

Davidson, R. and MacKinnon, J. G. (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York.

Diebold, F. X. (2004), *Elements of Forecasting*, third edn, Thomson South-Western, Mason, Ohio.

Diebold, F. X. and Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics* **13**, 253–263.

Draper, N. R. and Smith, H., eds (1981), *Aplied Regression Analysis*, revised edn, John Wiley & Sons, New York.

Dufour, J.-M. (1980), 'Dummy variables and predictive tests for structural change', *Economics Letters* **6**, 241–247.

Dufour, J.-M. (1997), 'Some impossibility theorems in econometrics, with applications to structural and dynamic models', *Econometrica* **65**, 1365–1389.

Dufour, J.-M. (2003), 'Identification, weak instruments and statistical inference in econometrics', *Canadian Journal of Economics* **36**(4), 767–808.

Dufour, J.-M. and Ghysels, E., eds (1996), *Recent Developments in the Econometrics of Structural Change*, Vol. 70 (No. 1) of *Journal of Econometrics (Annals Issue)*, North-Holland, Amsterdam. 316 pages.

Dufour, J.-M., Ghysels, E. and Hall, A. (1994), 'Generalized predictive tests and structural change analysis in econometrics', *International Economic Review* **35**, 199–229.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2003), *Bayesian Data Analysis*, second edn, Chapman and Hall/CRC.

Godfrey, L. G. (1988), *Misspecification Tests in Econometrics : The Lagrange Multiplier Principle and Other Approaches*, Cambridge University Press, Cambridge, UK.

Gouriéroux, C. and Monfort, A. (1995), *Statistics and Econometric Models, Volumes One and Two*, Cambridge University Press, Cambridge, U.K.

Grasa, A. A. (1989), *Econometric Model Selection: A New Approach*, Vol. 16 of *Advanced Studies in Theoretical and Applied Econometrics*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Hannan, E. J. and Quinn, B. (1979), 'The determination of the order of an autoregression', *Journal of the Royal Statistical Society, Series B* **41**, 190–191.

Harvey, D. I., Leybourne, S. J. and Newbold, P. (1997), 'Testing the equality of prediction mean squared errors', *International Journal of Forecasting* **13**, 281–291.

Hocking, R. R. (1976), 'The analysis and selection of variables in linear regression', *Biometrika* **32**, 1–49.

Hurvitch, C. M. and Tsai, C.-L. (1989), 'Regression and time series model selection in small samples', *Biometrika* **76**, 297–307.

Judge, G. G. and Bock, M. E. (1978), *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*, North-Holland, Amsterdam.

Judge, G. G., Griffiths, W. E., Carter Hill, R., Lütkepohl, H. and Lee, T.-C. (1985), *The Theory and Practice of Econometrics*, second edn, John Wiley & Sons, New York.

Keuzenkamp, H. A. (2000), *Probability, Econometrics and Truth: The Methodology of Econometrics*, Cambridge University Press, Cambridge, U.K.

Krishnaiah, P. R. and Kanal, L. N., eds (1982), *Handbook of Statistics 2: Classification, Pattern Recognition and Reduction of Dimensionality*, North-Holland, Amsterdam.

Lancaster, T. (2004), *An Introduction to Modern Bayesian Econometrics*, Blackwell, Oxford, U.K.

Le Cam, L. (1953), 'On some asymptotic properties of maximum likelihood estimates and related Bayes estimates', *University of California Publications in Statistics* **1**, 277–330.

Leamer, E. (1978), *Specification Searches: Ad Hoc Inferences with Nonexperimental Data*, John Wiley & Sons, New York.

Leamer, E. E. (1983), Model choice, *in* Z. Griliches and M. D. Intriligator, eds, 'Handbook of Econometrics, Volume 1', North-Holland, Amsterdam, chapter 5, pp. 285–329.

Leeb, H. and Pötscher, B. (2005), 'Model selection and inference: Facts and fiction', *Econometric Theory* **21**, 29–59.

Lehmann, E. L. (1986), *Testing Statistical Hypotheses*, 2nd edn, John Wiley & Sons, New York.

MacKinnon, J. G. (1992), 'Model specification tests and artificial regressions', *Journal of Economic Litterature* **30**, 102–46.

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

Mariano, R. S. (2002), Testing forecast accuracy, *in* (Clements and Hendry 2002), chapter 13, pp. 284–298.

McCracken, M. W. and West, K. D. (2002), Inference about predictive ability, *in* (Clements and Hendry 2002), chapter 14, pp. 299–321.

McQuarrie, A. D. R. and Tsai, C.-L. (1998), *Regression and Time Series Model Selection*, World Scientific, Singapore.

Meese, R. A. and Rogoff, K. (1988), 'Was it real? The exchange rate-interest differential relation over the modern floating -rate period', *Journal of Finance* **43**, 933–948.

Miller, A. (2002), *Subset Selection in Regression*, second edn, Chapman & Hall / CRC, Boca Raton, Florida.

Morgan, M. S. and Morrison, M., eds (1999), *Models as Mediators: Perspectives on Natural and Social Science*, Cambridge University Press, Cambridge, U.K.

Parzen, E. (1977), Multiple time series: Determining the order of approximating autoregressive schemes, *in* P. R. Krishnaiah, ed., 'Multivariate Analysis IV: Proceedings of the Fourth International Symposium on Multivariate Analysis', North-Holland/Elsevier, Amsterdam.

Pesaran, M. H., Smith, R. P. and Yeo, J. S. (1985), 'Testing for structural stability and predictive failure: A review', *Manchester School* **3**, 280–295.

Poirier, D. J. (1994), *The Methodology of Econometrics, 2 Volumes*, Edward Elgar, Aldershott, England.

Popper, K. (1968), *The Logic of Scientific Discovery*, revised edn, Harper Torchbooks, New York.

Pötscher, B. (1991), 'Efffects of model selection on inference', *Econometric Theory* **7**, 163–185.

Pötscher, B. (2002), 'Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots and estimation of long memory parameters', *Econometrica* **70**(3), 1035–1065.

Pötscher, B. M. (1983), 'Order estimation in ARMA-models by Lagrange multiplier tests', *Annals of Statistics* **11**, 872–885.

Sakamoto, Y., Ishiguro, M. and Kitagawa, G. (1985), *Akaike Information Criterion Statistics*, Reidel, Dordrecht, The Netherlands.

Sawa, T. (1978), 'Information criteria for discriminating among alternative regression models', *Econometrica* **46**, 1273–1291.

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

Shibata, R. (1976), 'Selection of the order of an autoregressive model by Akaike's information criterion', *Biometrika* **71**, 117–126.

Shibata, R. (1980), 'Asymptotically efficient selection of the order of the model for estimating parameters of a linear process', *The Annals of Statistics* **8**, 147–164.

Theil, H. (1957), 'Specification errors and the estimation of economic relationships', *Review of the International Statistical Institute* **25**, 41–51.

Theil, H. (1961), *Economic Forecasts and Policy, 2nd Edition*, North-Holland, Amsterdam.

West, K. D. (1996), 'Asymptotic inference about predictive ability', *Econometrica* **64**, 1067–1084.

West, K. D. and McCracken, M. D. (1998), 'Regression-based tests of predictive ability', *International Economic Review* **39**, 817–840.

White, H. (2000), 'A reality check for data snooping', *Econometrica* **68**, 1097–1126.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons, New York.

Zellner, A., Keuzenkamp, H. A. and McAleer, M., eds (2001), *Simplicity, Inference and Modelling: Keeping It Sophisticatedly Simple*, Cambridge University Press, Cambridge, U.K.