FORTHCOMING: THE NEW PALGRAVE DICTIONARY OF ECONOMICS

K000069

Kernel estimators in econometrics

The kernel estimation method is a nonparametric procedure for analysing economic models. It is a data-based procedure which avoids the a priori parametric specification of the economic model, and it has become popular because of its wide applicability and well-developed theory. A substantial literature has developed where the local polynomial kernel estimator has been proposed to analyse various economic models, which include regression models, single-index models, dynamic time series models and panel data models. The frontier of this subject is expected to develop further in both theory and applications, especially with advances in computer technology.

For empirical research, we draw from economic theory the types of variables which can be used in the economic relationship (model) under consideration. But theory usually does not provide the functional form of the economic model. Empirical and theoretical work in econometrics is, therefore, often carried out by assuming linear or nonlinear parametric functional forms of the economic models (see Gallant, 1987, for work on nonlinear models by econometricians sparked by the work of statisticians Hartley, 1961, and Jennrich, 1969). However, these parametric models may often be mis-specified and hence they may provide biased and misleading conclusions. With this in view, econometrics moved in the direction of local modelling (local averaging), which is a data-based approach, for studying the economic relationships of unknown forms. In the regression framework this approach is also called 'nonparametric regression' or 'nonparametric smoothing'. Here our focus is on nonparametric kernel regression.

Nonparametric kernel regression methods are becoming increasingly popular for applied data analysis; they are best suited to situations involving large data-sets for which the number of variables involved is manageable. A kernel is simply a weighting function. The kernel estimation procedure was developed in the seminal published work of Rosenblatt (1956) on the density function, and later in the context of the regression function, by Nadaraya (1964) and Watson (1964). A detailed development on this subject in statistics was first presented by Prakasa Rao (1983), and then Härdle (1990) and Fan and Gijbels (1996), followed by the work of Pagan and Ullah (1999) in econometrics. There are other ways to do local modelling – for example, spline methods, series methods, differencing methods, and neural network methods (see Pagan and Ullah, 1999) – but the kernel smoothing procedure has become popular because of its vast applicability, simplicity, and well-developed theoretical underpinnings for both time-series and cross-section data. Nonparametric kernel methods essentially involve local averaging in a regression context: we can obtain a consistent estimate of the conditional mean by locally averaging those values of the dependent variable which are 'close' in terms of the values taken on by the regressors. The amount of local information used to construct the average is determined by a window width, also known as a 'bandwidth' or a 'smoothing parameter'.

Suppose one wished to estimate the function m in the regression equation:

$$y_i = m(x_i) + u_i,\ i = 1, \ldots, n \tag{1}$$

where y_i is the dependent variable, x_i is a vector of q regressors, and u_i is an additive error. A parametric approach intends to fit the data to a parametric model $m(x_i) = m(x_i, \theta)$, often a linear model with $m(x_i, \theta) = \alpha + x_i\beta$, where θ is a parameter set of the model. But from the perspective of economic theory many economic models tend to be nonlinear. This makes the linear model specification inappropriate for understanding economic relationships. A way to capture the nonlinearity in data is to model the regression function locally, that is, to obtain the regression function $m(x)$ at a given point x by applying the linear regression technique to the data in a window width of size h using the linear model

$$y_i = \alpha(x) + (x_i - x)\beta(x) + u_i, \text{ for } x_i \text{ in } x \pm \frac{h}{2} \tag{2}$$

This local linear regression method leads to the following locally weighted minimization problem:

$$\min \sum_{i=1}^{n} (y_i - \alpha(x) - (x_i - x)\beta(x))^2 \, K\left(\frac{x_i - x}{h}\right) \tag{3}$$

where $K(\cdot)$, a non-negative weight (kernel) function, is a decreasing function of distances of x_i from the point x, and h is a window width that determines how rapidly the weights decrease as the distance of x_i from x increases. Let $\hat{\alpha}(x)$ and $\hat{\beta}(x)$ be the estimated local linear least squares estimators, which are the solutions of (3). Then the estimated regression function at the point $x_i = x$ is $\hat{m}(x) = \hat{\alpha}(x)$ and $\hat{\beta}(x)$ is the estimator of $\beta(x) = \partial m(x)/\partial x$ which is the local slope. If $\beta(x) = 0$ in (2) and (3), then the resulting estimator of $m(x) = \alpha(x)$ is the Nadaraya (1964) and Watson (1964) kernel regressor estimator. The local linear regression approach in (2) amounts to considering a linear Taylor series expansion of $m(x_i)$ around x in model (1). This approach can be extended to a local polynomial regression by taking a polynomial expansion of order, say p, of $m(x_i)$ around x. This provides the local polynomial least squares estimator of $m(x)$ (Stone, 1977). The local linear estimators ($p = 1$) perform better than the Nadaraya–Watson estimator ($p = 0$) with respect to bias reduction, absence of boundary effects, and the adaptation to various design situations; for $p \geq 1$ the local polynomial estimators may suffer singularity problems in applied settings.

The principle of local regression estimators can be generalized to other parametric regression settings such as local logit and probit, local proportional hazards, local quantile, robust regression, and nonlinear time-series models. For example, if we let $m(x_i,\theta)$ be a parametric model and $L_i(y_i, x_i, m(x_i, \theta))$ be the loss or the log-likelihood of the i-th observation, then we can minimize (if a loss) or maximize (if a likelihood) the objective function given by

$$L(\theta) = \sum_{i=1}^{n} L_i(y_i, x_i, m(x_i, \theta)) \, K\left(\frac{x_i - x}{h}\right) \tag{4}$$

The $m(x_i, \theta)$ is now locally estimated by $m(x_i, \hat{\theta}(x))$, for example, when $m(x_i, \theta) = \alpha + x_i\beta$ then $m(x_i, \theta(x)) = \alpha(x) + \beta(x) \quad x_i$ and $L_i(y_i, x_i, m(x_i, \theta(x)) = (y_i - \alpha(x) - (x_i - x)\,\beta(x))^2$, or $L(\theta)$ is maximized with L_i written presuming normality of errors. Similarly, in a single index econometric model $L_i(y_i, x_i, m(x_i, \theta)) = \log[F(x_i\beta)^{y_i}(1 - F(x_i\beta))^{1-y_i}]$ where $y_i = 1$ or 0 and $F(\cdot)$ is a cumulative distribution function, and in the case of a local linear $k\text{−}th$ quantile regression $L_i = u_i(k - I(u_i < 0))$ where $u_i = y_i - \alpha(x) - (x_i - x)\,\beta(x), 0 < k < 1$, and $I(\cdot)$ is the usual indicator function.

The selection of window width h is by far the most important issue of nonparametric kernel estimation. When h is arbitrarily small, the bias of the estimator is small but the variance is large. Conversely, when h is large, the estimator has a lower variance and a higher bias. Much of the literature on the methods of window width selection can really be viewed as attempts to balance this classic bias–variance trade-off. Overall, the selection rules fall into roughly three broad categories: (*a*) reference rules that would be optimal from a reference data generating process, (*b*) plug-in, penalizing, and cross-validation methods, and (*c*) bootstrap methods (see Pagan and Ullah, 1999, and Marron, 1992).

The asymptotic properties of the local polynomial estimators are well established (see Fan and Gijbels, 1996, for cross-section data and Masry, 1996, for the time-series case). The implication of these results is that the rate of convergence of the pointwise estimator of the $r\!-\!th$ derivative of $m(x)$ is the inverse of $(nh^{q+2r})^{\frac{1}{2}}$, $r \geq 0$, which is slower than the parametric rate of $\sqrt{n}$. In fact, as the dimensions of the regressors q for a given r increase, the rates become worse, which is the well-known 'curse of dimensionality' problem. However, the rate of convergence of the average of the pointwise estimators (global estimators) of these derivatives is widely known to have $\sqrt{n}$ rate of convergence. One of the most popular ways to deal with the 'curse of dimensionality' is to consider the nonparametric additive regression model which can be written as $y_i = \beta_0 + \sum_{j=1}^{q} m(x_{ij}) + u_i$. Imposing this additivity provides an estimator having a one-dimensional nonparametric rate of convergence.

In recent years the kernel regression estimation methods have progressed in various directions. These include testing for the significance of a regressor or group of regressors, consistent testing for the correct parametric functional form, estimation of the so-called 'structural relationship' among endogenous (dependent) variables, and the estimation of various types of semiparametric models consisting of a combination of parametric and nonparametric models (see Pagan and Ullah, 1999). Extensive work on the empirical applications of the kernel regression estimation have begun to appear in both cross-section econometrics and time-series econometrics, especially in labour economics and empirical finance. Although some related work is being done, several challenging research issues remain to be worked out. The first is the development of a unified approach towards a data-driven window width, and the development of software that permits fast computation of kernel-based estimators and test statistics for large data-sets in a desktop environment. The second is the development of kernel-based estimation of time-series models for non-stationary data. Third is the systematic development of the work on kernel estimation of panel-data models with heterogeneity parameters, especially when the time-series component of the data is large. Finally, the development of the theory of kernel estimation of various econometric models with both continuous and discrete variables is important, especially for the empirical applications of the kernel regression methods (see Racine and Li, 2004).

The nonparametric kernel regression method is a dynamic area, and there are rapid ongoing theoretical advances. With advances in computer technology, applications of the kernel regression approach continue to increase. The developments described above provide the dimensions in which the kernel-estimation procedures have been explored in econometrics and statistics. In a broad sense, the frontier of this research area has moved on, and is expected to continue with further developments in both its theory and applications.

Aman Ullah

See also

<xref=E000007> econometrics;
<xref=N000152> nonlinear time series analysis;
<xref=xyyyyyy> nonparametric models and methods;
<xref=R000254> robust estimators in econometrics.

Bibliography

Fan, J. and Gijbels, I. 1996. *Local Polynomial Modeling and its Applications*. London: Chapman and Hall.

Gallant, A. 1987. *Nonlinear Statistical Models*. New York: John Wiley & Sons.

Härdle, W. 1990. *Applied Nonparametric Regression*. New York: Cambridge University Press.

Hartley, H. 1961. The modified Gaus–Newton method for the fitting of nonlinear regression functions by least squares. *Technomterics* 3, 269–80.

Jennrich, R. 1969. Asymptotic properties of nonlinear least squares estimators. *Annals of Mathematical Statistics* 40, 633–43.

Marron, S. 1992. Bootstrap bandwidth selection. In *Exploring the Limits of Bootstrap*, ed. R. Lepage and L. Billard. New York: John Wiley.

Masry, E. 1996. Multivariate regression estimation local polynomial fitting for time series. *Stochastic Process Application* 65, 81–101.

Nadaraya, E. 1964. On estimating regression. *Theory of Probability and its Application* 9, 141–2.

Pagan, A. and Ullah, A. 1999. *Nonparametric Econometrics*. New York: Cambridge University Press.

Prakasa Rao, B. 1983. *Nonparametric Functional Estimation*. New York: Academic Press.

Racine, J. and Li, Q. 2004. Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics* 119, 99–130.

Rosenblatt, M. 1956. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* 27, 642–69.

Stone, C. 1977. Consistent nonparametric regression. *Annals of Statistics* 5, 267–84.

Watson, G. 1964. Smooth regression analysis. *Sankhya, Series A* 26, 359–72.

Index terms

bootstrap
convergence
cross-section econometrics
curse of dimensionality
kernel estimators in econometrics
linear models
local modelling
nonlinear models
nonlinear time-series analysis
nonparametric regression
time-series econometrics

Index terms not found:

nonlinear models
nonlinear time-series analysis