Dantzig, G. 1963. *Linear Programming and Extensions.* Princeton: Princeton University Press.
Dennis, J.E. 1984. A user's guide to nonlinear optimization algorithms. *Proceedings IEEE* 72, 1765–76.
Dorfman, R., Samuelson, P.A. and Solow, R.M. 1958. *Linear Programming and Economic Analysis.* New York: McGraw-Hill.
Gale, D. 1960. *The Theory of Linear Economic Models.* New York: McGraw-Hill.
Gass, S.I. 1975. *Linear Programming: Methods and Applications.* 4th edn. New York: McGraw-Hill.
Hadley, G. 1963. *Linear Programming.* Reading, MA: Addison-Wesley.
Hadley, G. 1964. *Nonlinear and Dynamic Programming.* Reading, MA: Addison-Wesley.
Hestenes, M.R. 1975. *Optimization Theory: The Finite Dimensional Case.* New York: Wiley.
Hicks, J.R. 1946. *Value and Capital.* 2nd edn. New York: Oxford University Press.
Intriligator, M.D. 1971. *Mathematical Optimization and Economic Theory.* Englewood Cliffs, NJ: Prentice-Hall.
Intriligator, M.D. 1981. Mathematical programming with applications to economics. Chapter 2 in *Handbook of Mathematical Economics*, vol. 1, ed. K.J. Arrow and M.D. Intriligator. Amsterdam: North-Holland.
John, F. 1948. Extremum problems with inequalities as side conditions. In *Studies and Essays: Courant Anniversary Volume,* ed. K.O. Friedrichs, O.W. Neugebauer, and J.J. Stoker. New York: Interscience Publishers.
Kuhn, H.W. and Tucker, A.W. 1951. Nonlinear Programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability,* ed. J. Neyman. Berkeley: University of California Press.
Lancaster, K. 1968. *Mathematical Economics.* New York: Macmillan.
Luenberger, D.G. 1973. *Introduction to Linear and Nonlinear Programming.* Reading, MA: Addison-Wesley.
Mangasarian, O.L. 1969. *Nonlinear Programming.* New York: McGraw-Hill.
Martos, B. 1975. *Nonlinear Programming: Theory and Methods.* Amsterdam: North-Holland.
McCormick, G.P. 1983. *Nonlinear Programming: Theory, Algorithms, and Applications*. New York: Wiley.
Nadiri, M.I. 1982. Producers theory. Chapter 10 in *Handbook of Mathematical Economics*, vol. 2, ed. K.J. Arrow and M.D. Intriligator. Amsterdam: North-Holland.
Phlips, L. 1983. *Applied Consumption Analysis.* Revised edn. Amsterdam: North-Holland.
Polak, E. 1971. *Computational Methods in Optimization: a Unified Approach*. New York: Academic Press.
Samuelson, P.A. 1947. *Foundations of Economic Analysis.* Cambridge, MA: Harvard University Press.
Schittkowski, K. 1980. *Nonlinear Programming Codes.* Berlin: Springer-Verlag.
Takayama, A. 1985. *Mathematical Economics.* 2nd edn. New York: Cambridge University Press.
Wold, H. and Jureen, L. 1953. *Demand Analysis.* New York: Wiley.
Zangwill, W.I. 1969. *Nonlinear Programming: A Unified Approach.* Englewood Cliffs, NJ: Prentice-Hall.

nonlinear time series analysis

Since the early 1980s, there has been a growing interest in stochastic nonlinear dynamical systems of the form

$$x_{t+1} = f(x_t, x_{t-1}, \ldots, x_{t-p}) + \sigma(x_t)\varepsilon_t, \tag{1}$$

where $\{x_t\}_{t=0}^{\infty}$ is a zero mean, covariance stationary process, $f : R^{p+1} \to R$, σ is the conditional volatility, and $\{\varepsilon_t\}_{t=0}^{\infty}$ is an independent and identically distributed noise process. The major recent developments in nonlinear time series are described here using this canonical model. The first section develops representation theory for a third order approximation. Nonparametric approaches follow; these rely on series expansions of the general model. Ergodic properties including path dependence and dimension are considered next. I then consider two widely utilized parametric models, piecewise linear models of f and autoregressive models for volatility. I conclude with a discussion of hypothesis testing and forecasting.

Volterra expansion

There is no general causal representation for nonlinear time series as in the linear case. Series approximations rely on the *Volterra expansion*,

$$\begin{aligned} x_{t+1} \simeq f(0) &+ \sum_{i=1}^{p} f_{i_1} x_{t-i_1} + \sum_{i_1=1}^{p}\sum_{i_2=i_1}^{p} f_{i_1 i_2} x_{t-i_1} x_{t-i_2} \\ &+ \sum_{i_1=1}^{p}\sum_{i_2=i_1}^{p}\sum_{i_3=i_2}^{p} f_{i_1 i_2 i_3} x_{t-i_1} x_{t-i_2} x_{t-i_3} + \cdots \end{aligned} \tag{2}$$

Brockett (1976) shows any continuous map over $[0, T]$ can be approximated by a finite Volterra series. Mittnik and Mizrach (1992) examine forecasts using generalized polynomial expansions like (2). Potter (2000) shows that in the cubic case, a one-sided Wold-type representation in terms of white noise v_t can be obtained,

$$\begin{aligned} x_{t+1} \simeq &\sum_{i=1}^{\infty} g_{i_1} v_{t-i_1} + \sum_{i_1=1}^{\infty}\sum_{i_2=i_1}^{\infty} g_{i_1 i_2} v_{t-i_1} v_{t-i_2} \\ &+ \sum_{i_1=1}^{\infty}\sum_{i_2=i_1}^{\infty}\sum_{i_3=i_2}^{\infty} g_{i_1 i_2 i_3} x_{t-i_1} x_{t-i_2} x_{t-i_3}. \end{aligned} \tag{3}$$

Koop, Pesaran and Potter (1996) note that the impulse response functions, $E[x_{t+n}|x_t, v_t] - E[x_{t+n}|x_t]$ will depend upon the size and sign of v_t as well as the current state x_t.

I now turn to nonparametric approaches which build on approximations like 2.

Nonparametric estimation

Consider the local polynomial approximation to $f(.)$ around x_0,

$$\widehat{f}(x) = \sum_{j=0}^{m} \beta_j (x - x_0)^j. \tag{4}$$

In the case $j = 0$, this corresponds to the *kernel regression* estimator of Nadaraya and Watson,

$$\widehat{f}(x) = \frac{\sum_{t=1}^{T} x_{t+1} K_h(x_t - x_0)}{\sum_{t=1}^{T} K_h(x_t - x_0)}. \tag{5}$$

The K_h are *kernels*, usually functions with a support on a compact set, assigning greater weight to observations closer to x_0. h is the *bandwidth* parameter, determining the size of the histogram bin. *Nearest neighbours* estimation is the case where h is adjusted to find a fixed number of nearby observations k.

More generally, the *local linear approximation* solves,

$$\alpha_0, \beta_0 \min \left(x_{t+1} - \alpha_0 - \beta_0 (x_t - x_0)\right)^2 \\ K_h(x_t - x_0). \tag{6}$$

The estimator (5) corresponds to the case where the only regressor in (6) is the constant term.

The application of these methods in the time series case is a fairly recent development. Conditions for consistency and asymptotic normality rely on *mixing conditions* where the dependence between x_{t+j} and x_t becomes negligible as j grows large.

A closely related approach involves the use of a *recurrent neural networks*,

$$\Psi_i(x_t, h_{t-1}) = \Psi\left(\gamma_{i0} + \gamma_{i1} x_t + \sum_{k=1}^{r} \delta_{ik} h_{k,t-k}\right),$$

$$x_{t+1} = \Phi\left(\beta_0 + \sum_{i=1}^{p} \beta_i \Psi_i(x_t, h_{t-1})\right). \tag{7}$$

Kuan, Hornik, and White (1994) provide convergence results for bounded Ψ (most commonly the logistic) as p grows large.

A popular approach in the frequency domain is wavelets. The *discrete wavelet transform* is

$$x_{t+1} = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \gamma(j,k) \Psi_{j,k}(t), \tag{8}$$

where the *mother wavelet* $\Psi(t)$,

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}} \Psi\left(\frac{t - k\tau_0 s_0^j}{s_0^j}\right), \tag{9}$$

is parameterized by scale s_0 and translation τ, and the wavelet coefficients are given by

$$\gamma(j,k) = \left\langle \Psi_{j,k}(t), x(t) \right\rangle. \tag{10}$$

Daubechies (1992) orthonormal basis functions,

$$E[\Psi_{j,k}(t)\Psi_{m,n}(t)] = 0, \quad \forall j \neq m, k \neq n, \tag{11}$$

have received the widest application.

Even when very little is known about f or σ, nonlinear time series analysis can shed light on the long run average or *ergodic* properties of the dynamical system.

Ergodic properties

Mathematicians have known since Poincaré that even simple maps like (1) can produce very complex dynamics. The nonlinear time series literature has developed tools for estimation of ergodic properties of these systems. Denote by $Df(\bar{x})$ the Jacobian matrix of partial derivatives of (1),

$$\begin{bmatrix} \partial f_1/\partial x_1 & \cdots & \partial f_1/\partial x_p \\ \vdots & \ddots & \vdots \\ \partial f_p/\partial x_1 & \cdots & \partial f_p/\partial x_p \end{bmatrix} \tag{12}$$

evaluated at $\bar{x}$. Replacing 12 with a sample analog,

$$J_t = \begin{bmatrix} \Delta f_1/\Delta x_{1,t} & \cdots & \Delta f_1/\Delta x_{p,t} \\ \vdots & \ddots & \vdots \\ \Delta f_p/\Delta x_{1,t} & \cdots & \Delta f_p/\Delta x_{p,t} \end{bmatrix} \tag{13}$$

we compute eigenvalues V_i,

$$V_i(Q_T' Q_T) \tag{14}$$

rank ordered from $1, \ldots, p$, where

$$Q_T = J_{T-p} \cdot J_{T-p-1} \cdots J_1 \tag{15}$$

The *Lyapunov exponents* are defined for the positive eigenvalues V_i^+ as

$$T \to \infty \lim \lambda_i = \frac{1}{2(T-p)} \ln V_i^+, \tag{16}$$

and a single exponent greater than 1 characterizes a system with sensitive dependence. Popularly known as 'chaos', this property implies that dynamic trajectories become unpredictable even when the state of the system is known with certainty. Gençay and Dechert (1992) and Shintani and Linton (2004) provide methods for estimating these. Shintani and Linton (2003; 2004) reject the presence of positive Lyapunov exponents in both real output and stock returns.

The sum of the Lyapunov exponents also provides a measure of the Kolmogorov–Sinai *entropy* of the system. This tells the researcher how quickly trajectories separate. Mayfield and Mizrach (1991) estimate this time at about 15 minutes for the S&P 500 index.

A final quantity of interest is the *dimension* p of the dynamical system. Nonlinear econometricians try to estimate the dimension from a scalar m-history. A powerful result due to Takens (1981) says this can be done as long as $m \geq 2p + 1$. Diks (2004) has shown that the scaling of correlation exponents seems to be consistent with the stochastic volatility model.

A great deal of progress has been made with parametric models of (1) as well. I begin with the widely utilized piecewise linear models.

Piecewise linear models

The most widely applied parametric nonlinear time series specification has been the Markov switching model introduced by James Hamilton (1989). The function f is a piecewise linear function,

$$f(x_t) = \left\{ \begin{array}{l} \mu^{(1)} + \sum_{j=0}^{p} \phi_j^{(1)}(x_{t-j} - \mu^{(1)}, S_t = s_t^{(1)}) \\ \vdots \\ \mu^{(m)} + \sum_{j=0}^{p} \phi_j^{(m)}(x_t - \mu^{(m)}, S_t = s_t^{(m)}) \end{array} \right\}, \tag{17}$$

where the changes among states are governed by an unobservable regime switching process, $S_t = s_t^{(i)}$, $i = 1, \ldots m$, an $m \times m$ transition matrix Π, and $E[x_t | S_t = s_t^{(i)}] = \mu^{(i)}$. When S_t is unobserved, $Pr(S_t | x_{t-1})$ is nonlinear in x_{t-1}. Hamilton has shown that a two-dimensional switching model describes well the business cycle dynamics in the United States. This model has been extended to include regime dependence in volatility (Kim, 1994) and time varying transition probabilities (Filardo, 1994).

The latent state vector requires forming prior and posterior estimates of which regime you are in. The EM algorithm (Hamilton, 1990) and Bayesian Gibbs sampling methods (Albert and Chib, 1993) have proven fruitful in handling this problem. Hypothesis testing is also nonstandard because under the alternative of $m - 1$ regimes, the conditional mean parameters are nuisance parameters. Hansen (1996) has explored carefully these issues.

A closely related framework is the *threshold autoregressive* (TAR) model,

$$f(x_t) = \left\{ \begin{array}{l} \left[\mu^{(1)} + \sum_{j=0}^{p} \phi_j^{(1)}(x_{t-j} - \mu^{(1)})\right] I(q(x_{t-d}, Z_t) \leq \gamma_1) \\ \left[\mu^{(2)} + \sum_{j=0}^{p} \phi_j^{(2)}(x_{t-j} - \mu^{(2)})\right] I(\gamma_1 < q(x_{t-d}, Z_t) \leq \gamma_2) \\ \vdots \\ \left[\mu^{(m)} + \sum_{j=0}^{p} \phi_j^{(m)}(x_t - \mu^{(m)})\right] I(q(x_{t-d}, Z_t) > \gamma_{m-1}) \end{array} \right\}. \tag{18}$$

$I(.)$ is the indicator function, and $q(x_{t-d}, Z_t)$, the regime switching variable, is assumed to be an observable function of exogenous variables Z_t and lagged x's. The integer d is known as the *delay parameter*. When q depends only upon x, the model is called *self-exciting*.

Teräsvirta (1994) has developed a two-regime version of the TAR model in which regime changes are governed by a smooth transition function $G(x_{t-d}, Z_t) : R^k \to [0, 1]$,

$$\begin{aligned} f(x_t) &= G(x_{t-d}, Z_t)_{j=0}^{p} \phi_j^{(1)}(x_{t-j} - \mu^{(1)}) \\ &\quad + (1 - G(x_{t-d}, Z_t))_{j=0}^{p} \phi_j^{(2)}(x_{t-j} - \mu^{(2)}). \end{aligned} \tag{19}$$

Luukkonen, Saikkonen and Teräsvirta (1988) have shown that inference and hypothesis testing in this model is often much simpler than in the piecewise linear models. Van Dijk and Franses (1999) have extended this model to multiple regimes. Applications of this framework have been widespread from macroeconomics (Teräsvirta and Anderson, 1992) to empirical finance (Franses and van Dijk, 2000).

Krolzig (1997) considers the multivariate case where $x_t = (x_{1,t}, x_{2,t}, \ldots, x_{k,t})'$ is $k \times 1$. Balke and Fomby (1997) introduced threshold cointegration by incorporating error correction terms into the thresholds. Koop, Pesaran and Potter (1996) develop a bivariate model of US GDP and unemployment where the threshold depends upon the depth of the recession.

I now turn to models that introduce nonlinearity through the error term.

Models of volatility

Engle and Bollerslev have introduced the generalized autoregressive conditional heteroskedasticity (GARCH) model,

$$h_t = \alpha_0 + _{i=1}^{q} \alpha_i \sigma^2(x_{t-i}) \varepsilon_{t-i}^2 + _{i=1}^{p} \beta_i h_{t-i}, \tag{20}$$

where $h_t = E[(x_t - E[x_t | \Omega_{t-1}])^2 | \Omega_{t-1}]$ is the *conditional variance*. This is just a Box–Jenkins model in the squared residuals of 1 of order $(max[p, q], p)$. The model is nonlinear because the disturbances are uncorrelated, but their squares are not.

The GARCH model describes the volatility clustering and heavy-tailed returns in financial market data, and has found wide application in asset pricing and risk management applications.

Volatility modelling has been motivated by the literature on options pricing. Popular alternatives to the GARCH model include the stochastic volatility (SV) model (Ghysels, Harvey and Renault, 1996), and the realized volatility approach of Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002). The discrete-time

SV model takes the form,

$$x_t = \sigma_\varepsilon \exp(h_t/2)\varepsilon_t, \tag{21}$$

$$h_t = \beta h_{t-1} + \sigma_h \eta_t,$$

where x_t is the demeaned log asset return, and ε_t and η_t are noise terms. Realized volatility sums high-frequency squared returns as an approximation of lower frequency volatility. Both GARCH and SV have been successful in explaining the departures from the Black–Scholes observed empirically.

The final two sections address the marginal contribution of nonlinear modelling to goodness of fit and forecasting.

Testing for linearity and Gaussianity

There is a large literature on testing the importance of the nonlinear components of a model. The most widely used test is due to Brock, Dechert, Scheinkman and LeBaron (BDSL, 1996). Their nonparametric procedure is built upon U-statistics. Serfling (1980) is a good introduction.

The first step is to form m-histories of the data,

$$x_t^m = (x_t, x_{t+1}, \ldots, x_{t+m-1}), \tag{22}$$

with joint distribution $F(x_t^m)$. Introduce the kernel $h: R^m \times R^m \to R$,

$$h(x_t^m, x_s^m) = I(x_t^m, x_s^m, \varepsilon) \equiv I\left[\left\|x_t^m - x_s^m\right\| < \varepsilon\right], \tag{23}$$

where $I(.)$ is the indicator function. The *correlation integral* of Grassberger and Procaccia (1983),

$$C(m,\varepsilon) \equiv \int_X \int_X I(x_t^m, x_s^m, \varepsilon) dF(x_t^m) dF(x_s^m), \tag{24}$$

is the expected number of m-vectors in an ε neighbourhood. A U-statistic,

$$C(m,N,\varepsilon) \equiv \frac{2}{N(N-1)}{}_{t=1}^{N-1}{}_{s=t+1}^{N} I(X_t^m, X_s^m, \varepsilon), \tag{25}$$

is a consistent estimator of 24. BDSL demonstrate the asymptotic normality of the statistic

$$\sqrt{N}\frac{S(m,N,\varepsilon)}{\sqrt{Var[S(m,N,\varepsilon)]}}\sqrt{N}d\times \to N(0,1), \tag{26}$$

where

$$S(m,N,\varepsilon) = C(m,N,\varepsilon) - C(1,N,\varepsilon)^m. \tag{27}$$

There is a multi-dimensional extension due to Baek and Brock (1992). De Lima (1997) explores the use of the BDSL under moment condition failure.

There is a direct relationship between nonlinear and non-Gaussian time series. In the model (1), even if the disturbance term ε_t is normal, nonlinear transformations of Gaussian noise will make x_t non-Gaussian. Testing for Gaussianity is then an instrumental part of the nonlinear time series toolkit.

Hinich (1982) has developed testing in the time domain using the *bicorrelation*,

$$\gamma(r,s) = {}_{t=1}^{s} x_t x_{t+r} x_{t+s}/(N-s), \quad 0 \le r \le s, \tag{28}$$

and in the frequency domain using the *bispectrum*,

$$B(\omega_1, \omega_2) = {}_{r=-\infty}^{\infty}{}_{s=-\infty}^{\infty}\gamma(r,s)\exp\left[-i(\omega_1 r + \omega_2 s)\right]. \tag{29}$$

For a Gaussian time series, the bicorrelation should be close to zero, and the bispectrum should be flat across all frequencies. Both tests have good power against skewed alternatives.

Ramsey and Rothman (1996) have proposed a related time domain procedure that looks for *time reversibility*,

$$\begin{aligned} &F(X_t, X_{t+1}, \ldots, X_{t+r}) \\ &\quad = F(X_{s-t}, X_{s-t-1}, \ldots, X_{s-t-r}) \end{aligned} \tag{30}$$

for any r, s and t, where $F(.)$ is the joint distribution. This condition is stronger than stationarity because of the triple index. The authors find evidence of business cycle asymmetry using this diagnostic.

Forecasting

For many, the bottom line on nonlinear modelling is the ability to generate superior forecasts. In this respect, the results from the nonlinear literature are decidedly mixed. Harding and Pagan (2002) are prominent sceptics. Teräsvirta, van Dijk and Medeiros (2005) provide a very wide set of evidence in favour of nonlinear models.

Aside from the comparison of point forecasts from model i, $u_{i,t+1} = x_{t+1} - f_i(x_t)$, with a particular loss function $g(.)$,

$$H_0 : E[g(u_{i,t+1}) - g(u_{j,t+1})] = 0, \tag{31}$$

there has been growing interest in comparing forecast densities $p_i(x_{t+1}|f_i(x_t))$,

$$H_0 : \int [p_i(x_{t+1}|f_i(x_t)) - p_j(x_{t+1}|f_j(x_t))]dx = 0. \tag{32}$$

Corradi and Swanson (2005) provide a comprehensive overview of available tools.

BRUCE MIZRACH

See also **forecasting; linear models; stochastic volatility models.**

I would like to thank Cees Diks, James Hamilton, Sebastiano Manzan, Simon Potter, Phil Rothman, Dick van Dijk and Steven Durlauf for helpful comments.

Bibliography

Albert, J. and Chib, S. 1993. Bayesian analysis via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics* 11, 1–15.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. 2003. Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.

Barndorff-Nielsen, O.E. and Shephard, N. 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society* , Series B 64, 253–80.

Baek, B. and Brock, W.A. 1992. A nonparametric test for independence of a multivariate time series. *Statistica Sinica* 2, 137–56.

Balke, N. and Fomby, T. 1997. Threshold cointegration. *International Economic Review* 38, 627–45.

Brock, W.A., Dechert, W.D., Scheinkman, J.A. and LeBaron, B. 1996. A test for independence based on the correlation dimension. *Econometric Reviews* 15, 197–235.

Brockett, R.W. 1976. Volterra series and geometric control theory. *Automatica* 12, 167–76.

Corradi, V. and Swanson, N.R. 2005. Predictive density evaluation. In *Handbook of Economic Forecasting*, ed. C.W.J. Granger, A. Timmermann and G. Elliott. Amsterdam: North-Holland.

Daubechies, I. 1992. *Ten Lectures on Wavelets*. 2nd ed. Philadelphia: SIAM.

De Lima, P. 1997. On the robustness of nonlinearity tests due to moment condition failure. *Journal of Econometrics* 76, 251–80.

Diks, C. 2004. The correlation dimension of returns with stochastic volatility. *Quantitative Finance* 4, 45–54.

Filardo, A.J. 1994. Business cycle phases and their transitional dynamics. *Journal of Business and Economic Statistics* 12, 299–308.

Franses, P.H. and van Dijk, D. 2000. *Nonlinear Time Series Models in Empirical Finance*. New York: Cambridge University Press.

Gallant, A.R. and Tauchen, G. 1987. Seminonparametric maximum likelihood estimation. *Econometrica* 55, 363–90.

Gallant, A.R. and Tauchen, G. 1996. Which moments to match? *Econometric Theory* 12, 657–81.

Gençay, R. and Dechert, W.D. 1992. An algorithm for the n Lyapunov exponents of an n-dimensional unknown dynamical system. *Physica D* 59, 142–57.

Ghysels, E., Harvey, A. and Renault, E. 1996. Stochastic volatility. In *Handbook of Statistics* 14: *Statistical Methods in Finance*, ed. G.S. Maddala and C.R. Rao. Amsterdam: North-Holland.

Granger, C.W.J. and Hallman, J. 1991. Nonlinear transformations of integrated time series. *Journal of Time Series Analysis* 12, 207–24.

Grassberger, P. and Procaccia, J. 1983. Measuring the strangeness of strange attractors. *Physica D* 9, 189–208.

Hamilton, J.D. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–84.

Hamilton, J.D. 1990. Analysis of time series subject to changes in regime. *Journal of Econometrics* 45, 39–70.

Hansen, B.E. 1996. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–30.

Harding, D. and Pagan, A. 2002. Dissecting the cycle: a methodological investigation. *Journal of Monetary Economics* 49, 365–81.

Hinich, M.J. 1982. Testing for Gaussianity and linearity of a stationary time series. *Journal of Time Series Analysis* 3, 169–76.

Kim, C.J. 1994. Dynamic linear models with Markov-switching. *Journal of Econometrics* 60, 1–22.

Koop, G., Pesaran, H. and Potter, S. 1996. Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics* 74, 119–48.

Krolzig, H.-M. 1997. *Markov Switching Vector Autoregressions: Modelling, Statistical Inference and Application to Business Cycle Analysis*. Berlin: Springer.

Kuan, C.-M., Hornik, K. and White, H. 1994. A convergence result for learning in recurrent neural networks. *Neural Computation* 6, 620–40.

Luukkonen, R., Saikkonen, P. and Teräsvirta, T. 1988. Testing linearity against smooth transition autoregressive models. *Biometrika* 75, 491–9.

Mayfield, S. and Mizrach, B. 1991. Nonparametric estimation of the correlation exponent. *Physical Review* A 88, 5298–301.

Mittnik, S. and Mizrach, B. 1992. Parametric and seminonparametric analysis of nonlinear time series. In *Advances in GLIM and Statistical Modeling*, ed. L. Fahrmeir, B. Francis, R. Gilchrist and G. Tutz. New York: Springer.

Potter, S. 2000. Nonlinear impulse response functions. *Journal of Economic Dynamics and Control* 24, 1425–46.

Ramsey, J.B. and Rothman, P. 1996. Time irreversibility and business cycle asymmetry. *Journal of Money, Credit, and Banking* 28, 1–21.

Serfling, R.J. 1980. *Approximation Theorems of Mathematical Statistics*. New York: John Wiley.

Shintani, M. and Linton, O. 2003. Is there chaos in the world economy? A nonparametric test using consistent standard errors. *International Economic Review* 44, 331–58.

Shintani, M. and Linton, O. 2004. Nonparametric neural network estimation of Lyapunov exponents and a direct test for chaos. *Journal of Econometrics* 120, 1–33.
Takens, F. 1981. Detecting strange attractors in turbulence. In *Springer Lecture Notes in Mathematics*, vol. 898, ed. D. Rand and L.-S. Young. Berlin: Springer.
Teräsvirta, T. 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89, 208–18.
Teräsvirta, T. and Anderson, H.M. 1992. Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics* 7, S119–S136.
Teräsvirta, T., van Dijk, D. and Medeiros, M.C. 2005. Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: a re-examination. *International Journal of Forecasting* 21, 755–74.
Van Dijk, D. and Franses, P.H. 1999. Modeling multiple regimes in the business cycle. *Macroeconomic Dynamics* 3, 311–40.

non-nested hypotheses

In economics, as in many other disciplines, there are competing explanations of the same phenomena, often characterized by alternative statistical models. Different models may represent, for example, different theoretical paradigms, or could be the result of alternative formulations from the same paradigm. Within the classical framework, the problem of model adequacy is approached through 'general specification tests', the 'diagnostic tests', and the 'non-nested tests'. All three approaches can be used to test the same explanation or hypothesis of interest (the null or the maintained hypothesis), but they differ in their consideration of the alternative(s). General specification tests intentionally consider a broad class of alternatives, while the alternatives considered under diagnostic and non-nested testing procedures are much more specific. In the case of non-nested tests the null hypothesis is contrasted to a specific alternative. Non-nested tests are appropriate when rival hypotheses are advanced for the explanation of the same economic phenomenon, and the aim is to devise a powerful test against a specific alternative.

When the null hypothesis is nested within the alternative, standard classical procedures such as those based on the likelihood ratio, Wald and Lagrange multiplier (or score) principles can be utilized. But if the null and the alternative hypotheses belong to 'separate' families of distributions, classical testing procedures cannot be applied directly and need to be suitably modified.

This article provides an overview of the concepts and some of the most widely used non-nested hypotheses tests and applies these procedures to the classical regression models. Our discussion of non-nested hypothesis testing will necessarily omit many topics. Survey articles on this subject include McAleer and Pesaran (1986), Gourieroux and Monfort (1994), and Pesaran and Weeks (2001).

Non-nested models

Suppose the object of interest is the process generating the random variable Y, observed over a sample of size n, $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$. Assume that the true process generating $\mathbf{y}$ is characterized by a joint probability density function, $f_0(\mathbf{y})$, which is unknown, and two models (hypotheses) are advanced as possible explanations of Y, represented by the joint probability density functions:

$$H_g = \{g(\mathbf{y}; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\},$$
$$H_h = \{h(\mathbf{y}; \boldsymbol{\gamma}), \boldsymbol{\gamma} \in \Gamma\}. \quad (1)$$

These functions are known but depend on a finite number of unknown parameters denoted by $\boldsymbol{\theta} \in \Theta$ and $\gamma \in \Gamma$, respectively. The sets Θ and Γ represent the 'admissible' parameter space for which the respective densities $g(\mathbf{y}; \boldsymbol{\theta})$ and $h(\mathbf{y}; \boldsymbol{\gamma})$ are well defined. The aim is to ascertain which of the two alternatives, H_g and H_h, if any, can be viewed as belonging to $f_0(\mathbf{y})$. In this set-up there is no natural null hypothesis; either of the two hypotheses under consideration can be taken as the null. In practice, the analysis of non-nested hypotheses is carried out with both alternatives taken in turn as the null hypothesis. Four outcomes are possible: (i) H_g rejected against H_h and not vice versa, (ii) H_h rejected against H_g and not vice versa, (iii) neither hypothesis is rejected against the other, and finally (iv) both hypotheses are rejected against one another. The first two outcomes are familiar from the classical test results and are straightforward to interpret. The third outcome can arise when the two models are very close to $f_0(\mathbf{y})$, and hence equivalent observationally. The fourth outcome suggests the existence of a third possible model which shares important features from both models under consideration.

Pseudo-true values and closeness measures

Given the observations $\mathbf{y}$, the maximum likelihood (ML) estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ are given by

$$\hat{\boldsymbol{\theta}}_n = \arg\max_{\theta \in \Theta} L_g(\boldsymbol{\theta}), \qquad \hat{\boldsymbol{\gamma}}_n = \arg\max_{\gamma \in \Gamma} L_h(\boldsymbol{\gamma}),$$

where the corresponding log-likelihood functions are defined by $L_g(\boldsymbol{\theta}) = \log(g(\mathbf{y}; \boldsymbol{\theta}))$ and $L_h(\boldsymbol{\gamma}) = \log(h(\mathbf{y}; \boldsymbol{\gamma}))$. Throughout we shall assume that probability densities satisfy the usual regularity conditions as established, for example in White (1982), such that $\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\gamma}}_n$ have asymptotically normal limiting distributions under the 'true' model, $f_0(\mathbf{y})$. In the general case where neither of the models under consideration coincide with $f_0(\mathbf{y})$, $\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\gamma}}_n$ are known as quasi-ML estimators and their