CEM

Center for Empirical Macroeconomics

Working Paper No. 40

Econometric Toolkit for Studying Dynamic Models in Economics and Finance

by

Peter Wöhrmann and Willi Semmler

University of Bielefeld
Department of Economics
Center for Empirical Macroeconomics
P.O. Box 100 131

33501 Bielefeld, Germany

Econometric Toolkit for Studying Dynamic Models in Economics and Finance

Peter Wöhrmann and Willi Semmler

January 2001

Abstract

The aim of this toolkit is to give students a practical introduction to linear econometric models as well as nonlinear econometric techniques that are widely applied in economics and finance. The particular purpose is to summarize tools and to collect computer programs that are useful in estimating dynamic relationships in economics and finance. The following topics are covered: (1) linear and nonlinear time series models and (2) estimation of intertemporal models.

Emphasizing the practical character of this toolkit, the computer programs of the estimated models are written in Gauss and available upon request.

Contents

Part I: Introduction

1 Basic Definitions

1.1 Random Variables, Distributions

Given a probability space $(\Omega, \mathcal{A}, P)$, a function $X : \Omega \to \mathbb{R}$ is called a continuous random variable it it is measurable, i.e.

$$\omega \in \Omega : X(\omega) \leq a \in \mathcal{A} \forall a \in \mathbb{R}.$$

Then the distribution and density function of X are defined by

$$F_X(x) : \mathbb{R} \to \mathbb{R} \forall x \in \mathbb{R}, F_X(x) = P_X((-\infty, x)) = P(\{\omega \in \Omega : X(\omega) < x\})$$

and

$$p_X(x) : \mathbb{R} \to \mathbb{R} \forall x \in \mathbb{R}, p_X(x) = F'_X(x),$$

respectively.

Appropriate measures to characterize a distribution are it's moments:

- mean and expected value

$$\mu_X = E(X) = \int_{-\infty}^{\infty} xp(x)dx,$$

- variance:

$$\sigma_X^2 = Var(x) = E((X - \mu_x)^2) = E(X^2) - \mu_X^2 \leq 0,$$

- skewness

$$E(X - \mu_X)^3,$$

  and

- kurtosis

$$E(X - \mu_X)^4.$$

Two random variables X_1 and X_2 are called independent if their joint and marginal distribution functions satisfy

$$F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2) \forall x_1, x_2 \in \mathbb{R}.$$

Then the following holds:

$$E(X_1X_2) = E(X_1)E(X_2),$$

i.e. they are not correlated and

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2)$$

Functions of Random Variables, $Y = f(X)$ are Random Variables, in particular the sum of two Gaussian random variables follows also a Gaussian distribution.

1.2 Convergence of Random Sequences

Often one is interested in the asymptotic behaviour of a sequence $X_1, X_2, ..., X_n, ...$ of random variables, i.e. the existence of a random variable X to which X_n for $n \to \infty$ converges.

Some useful convergence concepts

- Convergence with probability one (w.p.1)

$$P(\{\omega \in \Omega : \lim_{n\to\infty} \mid X_n(\omega) - X(\omega) \mid = 0\}) = 1$$

- Mean-square convergence

$$E(X_n^2) < \infty, \text{for } n = 1, 2, ..., E(X^2) < \infty \text{ and } \lim_{n\to\infty} E(\mid X_n - X \mid^2) = 0$$

- convergence in distribution

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

  at all continuity points of F_X.

Law of Large Numbers

Consider the sum of a sequence of independent identically distributed (i.i.d.) random variables $X_1, X_2, ...,$

$$A_n = \frac{1}{n} S_n = \frac{1}{n}(X_1 + X_2 + ... + X_n).$$

For $\mu = E(X_n)$ and $\sigma^2 = Var(X_n), E(A_n) = \mu$ and $Var(A_n) = \frac{\sigma^2}{2}$ follows from independence. The Law of Large Numbers proposes mean-square convergence as follows:

$$A_n \to \mu \text{ as } n \to \infty.$$

1.3 Stochastic Processes and Time Series

A sequence of random variables $X = \{X(t), t \in T\}$ is called a stochastic process if it describes the evolution of a stochastic system over time, i.e.

$$X : T \times \Omega \to \mathbb{R}$$

with $T = [0, \infty)$ (continuous time) or $T = \{t_1, t_2, ...t_n\}$ (discrete time).Then $X(\cdot, \omega) : T \to \mathbb{R}$ is a trajectory of the stochastic process. It is called stationary if

$$E(X_t) = \mu_X < \infty \text{ and } E(X_t - \mu_X)(X_\tau - \mu_X) = \sigma_{t\tau} < \infty.$$

The outcome of a stochastic process is called time series. The following models of stochastic processes are commonly used in economics and finance.

Markov Chains (temporal dependence)
A stochastic process X is called a Markov Chain, it it fulfills the condition

$$P(X_{n+1} = x_j \mid X_n = x_{i_n} = P(X_{n+1} = x_j \mid X_1 = x_{i_1}, ..., X_n = x_{i_n})$$

for all $x_j, x_{i_1}, ..., x_{i_n}$ in a given state space $\mathcal{X}$ and all $n = 1, 2, ...$

Wiener Processes (independence)
A standard Wiener Process $W = \{W(t), t \geq 0\}$ is defined to be a continuous Gaussian process with independent increments and $W(0) = 0$, w.p.1, $E(W(t)) = 0$ and $Var(W(t) - W(s)) = t - s$ $\forall 0 \leq s \leq t$. Then W is normally distributed and can be approximated in distribution on any finite time interval by means of a random walk.

Part II: Linear Time Series Models

2 Deterministic and Stochastic Trends

Trends, unlike the cyclical and irregular components have a permanent effect on a series. Most familiar trend model is a deterministic (polynomial) trend model:

$$y_t = a_0 + a_1 t + a_2 t^2 + ... + a_n t^n + \epsilon_t \tag{1}$$

as opposed to a stochastic trend model such as the random walk model:

$$y_t = y_{t-1} + \epsilon_t \tag{2}$$

The stochastic trend has been introduced by Nelson and Plotter (1982). The random walk model is just a special case of an AR(1),

$$y_t = a_0 + a_1 y_{t-1} + \epsilon_t \tag{3}$$

with $a_0 = 0$ and $a_1 = 1$.
The general solution to a random walk model is

$$y_t = y_0 + \sum_{i=1}^{t} \epsilon_i \tag{4}$$

which shows that the past shocks have a permanent, non-decaying effect on the series. In other words,

$$\frac{\partial y_t}{\partial \epsilon_t} = \frac{\partial y_t}{\partial \epsilon_{t-1}} = \frac{\partial y_t}{\partial \epsilon_{t-2}} = ... = \frac{\partial y_t}{\partial \epsilon_1} \tag{5}$$

The mean of the series is constant,

$$E[y_t] = y_0 \tag{6}$$

However, the variance is time dependent:

$$var(y_t) = var(\epsilon_t + \epsilon_{t-1} + ... + \epsilon_t) = t\sigma^2. \tag{7}$$

which indicates that the random walk process is non-stationary. Covariance of y_t and a_{t-s} is also time dependent:

$$\begin{aligned} E[y_t - y_0)(y_{t-s} - y_0)] &= [(\epsilon_t + \epsilon_{t-1} + ...\epsilon_t)(\epsilon_{t-s} + \epsilon_{t-s-1} + ...\epsilon_t)] \\ &= [(\epsilon_{t-s}^2 + \epsilon_{t-s-1}^2 + ...\epsilon_1^2)] \\ &= (t-s)\sigma^2 \end{aligned}$$

One should notice that the first difference of the series is stationary:

$$\Delta y_t = \epsilon_t \tag{8}$$

The correlation coefficient ρ_s is:

$$\begin{aligned}
\rho_s &= \frac{(t-s)\sigma^2}{\sqrt{t\sigma^2}\sqrt{(t-s)\sigma^2}} && (9)\\
&= \frac{(t-s)}{\sqrt{t(t-s)}} && (10)\\
&= \Big(\frac{t-s}{t}\Big)^{1/2} && (11)
\end{aligned}$$

Hence, the autocorrelation function (ACF) for a random walk process will show a slight tendency to decay. This also implies that ACF cannot distinguish between a unit root process ($a_1 = 1$) and processes with a_1 close to unity.

Graphically random walk processes are distinguished by the fact that they do not tend to return to a long-run (mean) value. The conditional mean of y_{t+1} is

$$E_t y_{t+1} = E_t(y_t + \epsilon_{t+1}) = y_t \tag{12}$$

Similarly, the conditional mean of y_{t+s} can be obtained from

$$y_{t+s} = y_t + \sum_{i=1}^{s} \epsilon_{t+i} \tag{13}$$

so that

$$\begin{aligned}
E_t y_{t+s} &= y_t + E_t\Big(\sum_{i=1}^{s} \epsilon_{t+i}\Big) && (14)\\
&= y_t && (15)
\end{aligned}$$

Hence, the constant value of y_t is the unbiased estimator of all future values of y_{t+s} for all $s > 0$.

2.1 Random Walk plus Drift Model

Random walk plus drift model is given by

$$y_t = y_{t-1} + a_0 + \epsilon_t \tag{16}$$

whose solution is

$$a_t = y_0 + a_0 t + \sum_{i=1}^{t} \epsilon_i \tag{17}$$

with a deterministic trend $a_0 t$ and a stochastic trend, $\sum_{i=1} \epsilon_i$. Again, notice that the first difference of the series yields a stationary series:

$$\Delta y_t = a_0 + \epsilon_t \tag{18}$$

The forecast function of random walk with drift model is different from the pure random walk model in that it is not flat but has a trend

$$\begin{aligned} y_{t+s} &= y_0 + a_0(t+s) + \sum_{i=1}^{t+s} \epsilon_i && (19) \\ &= y_t + a_0 s + \sum_{i=1}^{s} \epsilon_i && (20) \end{aligned}$$

and

$$E_t y_{t+s} = y_t + a_0 s \tag{21}$$

A more general model is obtained by augmenting the random walk plus model with a stationary noise process, $A(L)\eta_t$, which is called the general trend plus irregular model:

$$y_t = \mu_0 + a_0 t + \sum_{i=1}^{t} \epsilon_i + A(L)\eta_t \tag{22}$$

2.2 Removing the Trend

the appropriate form of eliminating the trend depends on the form of the trend, i.e., deterministic of stochastic, or both. If the nature of the trend is deterministic then detrending procedure s used where the series is regressed on a function of time and the residuals from that regression are regarded as the detrended series. However, if the trend is stochastic, then the appropriate procedure is to difference the series. Note, however, that differencing the time series will increase the volatility of the series.

2.2.1 Differencing

Take an $ARIMA(p, d, q)$:

$$A(L)y_t = B(L)\epsilon_t \tag{23}$$

Suppose that $A(L)$ has a single unit root and $B(L)$ has all roots outside the unit circle [i.e., the model is $(ARIMA(p,1,q)]$. Then, we can write

$$A(L) = (1-L)A*(L) \tag{24}$$

and

$$(1-L)A^*(L)y_t = B(L)\epsilon_t \tag{25}$$

or

$$A^*(L)\Delta y_t = B(L)\epsilon_t \tag{26}$$

The Δy_t series is stationary since all roots of $A^*(L)$ lie outside the unit circle. In general the d^{th} difference of a process with d unit roots is stationary. An $ARIMA(p,d,q)$ model has d unit roots; the d^{th} difference of such a model is a stationary $ARIMA(p,q)$ process. A series with d unit roots is said to be integrated of order d or simply $I(d)$.

2.2.2 Detrending by Filtering Out the Time Trend

Nonstationary processes with deterministic trends, however, cannot be transformed into well-behaved ARMA models by differencing. Consider,

$$y_t = a_0 + a_1 t + \epsilon_t \tag{27}$$

Differencing this process yields

$$\Delta y_t = a_1 + \epsilon_t - \epsilon_{t-1} \tag{28}$$

which is stationary but noninvertible. The appropriate way is to regress y_t on a constant and linear trend and retrieve the residual as the stationary component.

More generally a time series may have a polynomial time trend

$$y_1 = a_0 + a_1 t + a_2 t^2 + ... + a_n t^n + \epsilon_t$$

in which case detrending is accomplished by regressing y_t on a polynomial time trend. The appropriate degree of the polynomial can be determined by standard $t-$tests, $F-$tests and/or AIC or SBC.

2.2.3 Detrending by the Exponential Smoothing and HP-Filters

The HP-filter as proposed by Hodrick and Prescott (1980) separates the growth component (y^g = trend component) which can represent a (i) deterministic or a (ii) stochastic trend and y^c = cyclical component. We then have

$$y_t = y_t^g + y_t^c.$$

For a linear filter it holds:

$$y_t^g = \sum_{j=-\infty}^{\infty} g_j y_{t-j};$$

then $y_t^c = y_t - y_t^g$. The linear filter, as the HP- filter below, is a symmetric low frequency filter. The exponential smoothing filter reads

$$\min_{\{y_t\}_{t=1}^T} \sum_{t=1}^{T} (y_t - y_t^g)^2 + \lambda (y_t^g - y_{t-1}^g)^2.$$

The HP-filter is obtained by

$$\min_{\{y_t\}_{t=1}^T} \sum_{t=1}^{T} (y_t - y_t^g)^2 + \lambda[(y_{t+1}^g - y_t^g) - (y_t^g - y_{t-1})]^2.$$

with λ as a penalty parameter for changes in the trend (acceleration of growth) component. In practical applications one sets $\lambda = 1600$ for the HP-filter (see the Gauss program Hpfilt.g).[1] The HP-filter, and recently the Band Pass Filter, are frequently employed in macro dynamic modeling, and in the RBC literature, see Stock and Watson (1997). The Band Pass Filter also lists in Gauss.

2.2.4 Detrending by Other Smoothing Techniques

Detrending might also be undertaken by removing piecewise linear (segmented) trends or trends generated by moving averages, see the computer program by Franke and Semmler (1993) which is used in Franke and Semmler (1994).

2.2.5 Difference Versus Trend Stationary Models

A difference stationary (DS) model can be transformed into a stationary model by differencing and a trend stationary (TS) model can be transformed into a stationary model by removing the deterministic trend. At it was not appropriate to difference a TS series it is also inappropriate to detrend a DS series.

[1] See also the Band Pass Filter by Baxter and King (1995).

2.3 Testing for Trends and Unit-Roots

Since standard econometric theory depends on the stationarity assumption of the series and since the appropriate form of transforming a non-stationary series into a stationary one depends on the form of the trend we need a statistical devise, i.e. a test procedure, to determine the nature of the trend.

2.3.1 Unit Root Processes

Consider,

$$y_t = a_1 y_{t-1} + \epsilon_t \tag{29}$$

and suppose we want to test

$$H_0 : a_1 = 1 \tag{30}$$

Under the null hypothesis y_t is a random walk (a unit root process)

$$y_t = y_{t-1} + \epsilon_t \tag{31}$$

It has been shown that if one applies OLS to equation 29 to estimate a_1, where the actual $a_1 = 1$, one would obtain an estimate below unity. in other words OLS estimate of a_1 when it is equal to unity is downward biased. Furthermore, under the null hypothesis the $\hat{a}_1/se(\hat{a}_1)$ does not have the usual $t-$distribution. In other words we cannot use the critical values from the $t-$table to test hypothesis that $a_1 = 1$. **Unit Roots in a Regression Model**

Consider

$$y_t = a_0 + a_1 z_t + \epsilon_t \tag{32}$$

The assumptions of the classical regression model requires that both the y_t and z_t series are stationary and the errors have a zero mean and finite variance. However, if the series have unit roots (i.e. are non-stationary) we might have what is called a spurious regression. In such a case the regression has high R^2 and high $t-$statistics which are misleading because the customary statistical tests do not apply. The problem arises because normally the error process in a spurious regression it not stationary, i.e. contains a stochastic trend.

There are four cases with different consequences:

1. Both y_t and z_t are stationary $\Rightarrow$ the classical regression model is appropriate

2. y_t and z_t are integrated of different orders $\Rightarrow$ meaningless regression

3. y_t and z_t are integrated of the same order but the error term contains a stochastic trend $\Rightarrow$ spurious regression. If they are both $I(1)$ first difference formulation is appropriate
$$\Delta y_t = a_1 \Delta z_t + \Delta \epsilon_t \tag{33}$$

4. y_t and z_t are integrated of the same order but the error term is stationary $\Rightarrow$ the variables are co-integrated.

2.3.2 Dickey-Fuller Tests

An equivalent way of testing $H_0 : a_1 = 1$ in

$$y_t = a y_{t-1} + \epsilon_t \tag{34}$$

is to test $H_0 : \gamma = 0$ in

$$\Delta y_t = \gamma y_{t-1} + \epsilon_t \tag{35}$$

where $\gamma = a_1 - 1$. to test this hypothesis one estimates γ by applying OLS to the above equation and forms the $t-$statistic. The critical values for the test is provided by Dickey and Fuller which is now part of most of the econometric software (including Eviews). Those critical values however depend on whether an intercept and/or time trend is included in the regression equation. In other words there are three basic forms of regressions through which one can test for unit roots:

$$\Delta y_t = \gamma y_{t-1} + \epsilon_t \tag{36}$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + \epsilon_t \tag{37}$$

$$\Delta y_t = a_0 + a_2 t + \gamma y_{t-1} + \epsilon_t \tag{38}$$

One drawback of the Dickey-Fuller test is that it assumes that the errors are independent and have a constant variance. If the stochastic process generating the $y :_t$ series contains an ARMA component in addition to the stochastic and/or deterministic trend the residuals in equations 36, 37 and 38 will suffer from serial correlation and hence Dickey-Fuller test will not be appropriate. In such a case one applies Augmented Dickey-Fuller Test (ADF) which just ads an AR component to the above regressions and use the same critical values:

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{p} b_i \Delta y_{t-i} + \epsilon_t \tag{39}$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=1}^{p} b_i \Delta y_{t-i} + \epsilon_t \tag{40}$$

$$\Delta y_t = a_0 + a_2 t + \gamma y_{t-1} + \sum_{i=1}^{p} b_i \Delta y_{t-i} + \epsilon_t \tag{41}$$

The above procedure assumed that the null hypothesis is that there is only one unit root. If existence of two unit roots is suspected then the appropriate procedure is to first test for unit root in the first difference of the series and then if this is rejected test for the unit root in the level series.

3 Trends and "the Linear Model"

Considering the two-variable VAR (2) model

$$\begin{aligned} y_t &= a_0 + a_1 y_{t-1} + a_2 x_{t-1} + e_{1t} \\ x_t &= b_0 + b_1 x_{t-1} + b_2 y_{t-1} + e_{2t} \end{aligned}$$

with stationary time series x_t and y_t, ordinary least squares regression yields consistent and normally distributed estimates of a and b, if $X_t = (x_{t-1}, y_{t-1})$ is independent of the error terms e_{1t} and e_{2t}, respectively. In the case of non-stationary time series x_t and y_t with unit roots due to the higher convergence rates of moments of non-stationary processes than of moments of stationary processes, OLS-regression turns out to be a "superconsistent" estimate of a and b but the normal distribution of the parameters[2] can only be guaranteed if x_t and y_t are co-integrated, i.e. if their linear combination, the so called co-integration vector,[3]

$$z_t = y_t - \beta x_t$$

is stationary - thus co-integration of two non-stationary time series is characterized by similar trend components in both time series constituting a long-term dependence between them. Simple tests of co-integration can be performed using e.g. the Augmented Dickey-Fuller test of integration (see sect. 4) based on z_t.

4 Co-Integration and Error Correction Models

4.1 Derivation and Interpretation

Consider the two-variable (VAR (2) model

$$\begin{aligned} y_t &= a_0^1 + a_1^1 x_{t-1} + b_1^1 y_{t-1} + e_{1t} \\ x_t &= a_0^2 + a_1^2 y_{t-1} + b_1^2 x_{t-1} + e_{2t} \end{aligned}$$

[2]This property of parameter estimates is important to make statistical inferences about them.

[3]Note, that β refers to the OLS-regression of x_t on y_t.

Applying techniques of linear algebra, we restate this model in the following form,[4]

$$\Delta y_t = a_0^1 \Delta x_t - (1 - b_1^1)(y_{t-1} - (a_0^1 + a_1^1)x_{t-1} + e_{1t} \tag{42}$$
$$\Delta x_t = \underbrace{a_0^2 \Delta y_t}_{[1]} - \underbrace{(1 - b_1^2)}_{[2]} \underbrace{(x_{t-1} - (a_0^2 + a_1^2)y_{t-1}}_{[3]} + e_{2t} \tag{43}$$

called die Error Correction Model. The short-term dynamics [1] and long-term dependencies [2]&[3] can be analyzed simultaneously. Suppose, the error cannot become infinitely large, i.e. x_t and y_t are co-integrated, [3] can be interpreted as the deviation from the long-run equilibrium and [2] as it's correction.

4.2 Estimation and Tests

Suppose, x_t and y_t are co-integrated non-stationary time series with unit roots (see sect. 4.2.3), then parameters are estimated consistently using OLS-regression for each equation in the following reparameterization,

$$\Delta y_t = \bar{a}_0^1 \Delta x_t + \bar{b}_1^1 y_{t-1} + \bar{b}_2^1 x_{t-1} + e_{1t} \tag{44}$$
$$\Delta x_t = \bar{a}_0^2 \Delta y_t + \bar{b}_1^2 x_{t-1} + \bar{b}_2^2 y_{t-1} + e_{2t} \tag{45}$$

Alternatively one can employ the two-step procedure of Engle and Granger:

1. estimate the co-integration vector [3] of x_t and y_t (see sect. 4.2.3) and
2. substitute the co-integration vector in (42), (43) and estimate $a_0^{\{1,2\}}$ and $1 - b_1^{\{1,2\}}$.

This procedure is justified as the OLS-estimates of the parameters of stationary and non-stationary variables are asymptotically independent.
In the framework of (44), (45) Granger causality[5] tests can be designed in the following way: Δx_t Granger causes Δy_t, if $a_0^{-2} = b_2^{-2} = 0$ and $a_0^{-1} > 0$.
For applications of the ECM, see Hansen (1993).

[4] Note that $\Delta x_t = x_t - x_{t-1}$ and $\Delta y_t = y_t - y_{t-1}$, respectively.

[5] The concept of Granger causality is based on the proposition that the future does not influence the presence.

Part III: Nonlinear Time Series Models

In the last decade research on nonlinear models has found that economic and financial time series may be of high complexity (see especially chaos theory and related literature) which cannot possibly be generated by linear models. I our context, nonlinear dynamic models are defined as models where the dynamics of variables undergo regime changes as the variables vary. In general these are amplitude dependent models. For a survey of numerous nonlinear econometric models and for applications in economics and finance see the contributions in Semmler (1994).

5 Estimation of Continuous Time Models

Assume, for example, a second order nonlinear stochastic differential equation, such as the can der Pol equation

$$\ddot{x} - b(x)\dot{x} + bx = n \tag{46}$$

with

$$n : \text{white noise}$$
$$b(x) : a(1 - x^2)$$

with n a white noise term and a a coefficient. It can be estimated as follows. One can write eqn. (46) as a first order differential equation system in two variables such as

$$\dot{z} = f(z_t \mid \theta) + \omega_t \tag{47}$$

where θ is a parameter set and ω_t a white noise term. One can then solve the continuous time dynamic model (47)through the Euler method, Runge Kutta method, Milstein scheme or local linearizations[6]. Furthermore, one can directly estimate the parameters of the proposed model. We will discuss two estimation procedures.

5.1 The Euler Procedure

Employing, for example, the Euler procedure the discrete-time form of equation system (47) can be estimated by using actual data sets.
The Euler procedure amounts to estimate the parameter set θ in (47) using the following discretization,

[6]For a survey and comparison of the numerical accuracy of the different discretization methods, see Kloeden, Platen and Schurz (1991) and for the local linearization procedure, see Ozaki (1985, 1994).

$$z_{t+1} = z_t + hf(z_t \mid \theta) + \omega_t$$

where h is the step size (for examples, see Semmler and Kockesen (1997) and Chiarella, Semmler and Kockesen (1996)). Although the Euler procedure is a very convenient method of estimating a continuous time model, as has, however, been pointed out by Kloeden, Platen an Schurz (1991) and Ozaki (1994), it is not the most precise method of turning out stochastic differential equations into a discrete time estimable form. It is possible that instability can arise in the different equation although the corresponding differential equation is stable. This problem has been addressed in a series of papers by Ozaki. The Euler difference scheme is, however, still a useful approximation for estimating an equation system such as (47) in time discrete form by using Maximum Likelihood. Nonlinear Least Squares may also be a useful estimator, although it is not necessarily a consistent estimator for the above model. For the use of the Euler scheme in estimating a stochastic differential equation system such as (47 by ML, see Kloeden and Platen (1995: 242).

5.2 Local Linearization Procedure

In various papers Ozaki (1986, 1987, 1989, 1994) proposes a local linearization procedure that overcomes short comings of methods such as the Euler scheme. He suggests a local linearization of a nonlinear stochastic differential equation such as (47) by computing the Jacobian at each point in the state space, The nonlinear differential equation model (47) can be transformed into a discrete time model through linearization as follows (see Ozaki, 1994):

$$\begin{aligned}
z_{t+\Delta t} &= A(z_t)z_t + B(z_t)w_{t+\Delta z} \\
w_{t+\Delta z} &: \text{discrete time white noise} \\
A(z_t) &= \exp(L(z_t)\Delta t) \\
L(z_t) &= \frac{1}{\Delta t} log\{I + J_t^{-1}(e^{J_t \Delta t} - I)F_t\} \\
J_t &= \left\{\frac{\partial f(z)}{\partial z}\right\}_{z=z_t} \\
F_t &: \text{derived from } F_t z_t = f(z_t) \\
B(z_t) &: \text{function of the eigenvalues of } L(z_t)
\end{aligned}$$

This local linearization is consistent since $\Delta t \to 0$ the original differential equation is obtained. The estimation procedure can be undertaken by nonlinear least square or maximum likelihood procedures (see Ozaki 1994). The local linearization method of Ozaki is written in GAUSS. For an example, see Semmler and Kockesen (1997).

5.3 Estimation of a Random Walk

Random walks as stochastic processes are important tools in asset pricing, see Neftci (1996). A discrete time random walk can be expressed as

$$p_t = p_{t-1} + \epsilon_t, \epsilon_t = \begin{cases} \Delta & \text{with probability } \ \pi \\ -\Delta & \text{with probability } \ 1-\pi' \end{cases} \tag{48}$$

where ϵ_t is iid. (for details see Campbell, Lo and MacKinlay, 1996, ch. 9). A standard Brownian motion (Wiener process) which is frequently used in asset pricing is

$$P_t = \mu t + \sigma B_t \tag{49}$$

Conditional moments are

$$\begin{aligned} E[p_t \mid p_{t_0}] &= p_{t_0} + \mu(t - t_0) \\ Var[p_t \mid p_{t_0}] &= \sigma^2(t - t_0) \\ Cov[p_t \mid p_{t_0}] &= Var[p_{t_1}] = \sigma^2 t_1. \end{aligned}$$

In differential form we obtain from (49) a stochastic differential equation

$$dp_t = \mu dt + \sigma dB_t,$$

where μ is the drift and σB_t the diffusion term. Moreover,

$$dB_t \equiv \lim_{h \to dt} B_t + h - B_t.$$

The general form of the stochastic differential equation (SDE) with amplitude dependent coefficients is

$$dp_t = a(p, t, \alpha) + b(p, t, \beta) dB_t,$$

called a standard Wiener process with $\theta = \alpha, \beta)$ as unknown parameters. For a Wiener process with constant coefficients a ML estimation can be employed to estimate the parameter set θ.
Take

$$\begin{aligned} d\ log\ P &= \underbrace{(\mu - \frac{1}{2}\sigma^2)}_{\alpha} dt + \sigma dB \text{ and} \\ d\ log\ P &= \alpha dt + \sigma dB, \end{aligned}$$

where $\theta = (\alpha, \sigma)$ are constants. In discrete time form we have

$$log\ P_{t+h} = log\ P_t + \alpha h + \sigma dB_t,$$

where dB is white noise generated by a random number generator. For examples of such processes, see Neftci (1996, ch. 11).

As suggested by Campbell, Lo and MacKinlay (1996, ch. 9) we can employ the ML function

$$\mathcal{L}(\alpha, \sigma) = -\frac{1}{2} log(2\pi\sigma^2 h) - \frac{1}{2\sigma^2 h} \sum_{k=1}^{n} (r_k(h) - \alpha h)^2,$$

with r_k as continuously compounded returns and where h is now the observation interval of the data. Hence, estimates of α, σ can be written as follows:

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{nh} \sum_{k=1}^{n} (r_k(h) - \hat{\alpha} h)^2 \text{ and} \\ \hat{\alpha} &= \frac{1}{nh} \sum_{k=1}^{n} r_k(h). \end{aligned}$$

For the estimation of a more general SDE, the remarks in sect. 6.1 may be valid.

6 Threshold Regression Models

6.1 Regime Changes

The above van der Pol equation (46) can be approximated by a discrete time locally self-exciting but globally bounded system of the following type (see Ozaki 1985).

$$x_t = (\emptyset_1 + \pi e^{-x_{t-1}^2}) x_{t-1} + (\emptyset_2 + \pi_2 e^{-x_{t-1}^2}) x_{t-2} + \varepsilon_t$$

or equivalently by a (piecewise linear or nonlinear) threshold model such as

$$x_t = \begin{cases} \pi(T_1) x_{t-1} + \varepsilon_1 & \text{for } x_{t-1} < T_1 \\ \pi(x_{t-1}) x_{t-1} + \varepsilon_1 & \text{for } T_1 \leq x_{t-1} < T_2 \\ \pi(T_2) x_{t-1} + \varepsilon_1 & \text{for } x_{t-1} \geq T_2 \end{cases}$$

Threshold models have become popular since they offer a rich array of dynamic behavior and are able to identify models with amplitude-dependent behavior - see Tong (1990). See also the papers in Semmler (1994) which provide extensive reviews of the literature on these various frameworks.

6.2 Threshold Models

One particular class of time series models, called **Threshold Autoregressive (TAR)** models are based upon the principle of local approximations to a nonlinear system by introducing different linear regimes via thresholds. Tong (1990, p.99) refers to this principle as the threshold principle in the sense that as certain variables pass through thresholds the dynamic behavior of the system changes. In that sense they re suitable to analyze amplitude-dependence and regime changes and/or asymmetries and also amenable to examine nonlinear economic phenomena particularly within a multivariate context.

The univariate TAR models are called self-exciting thresholds autoregressive and moving average models (**SETARMA**). One particular case is the self-exciting threshold autoregressive (**SETAR**) model. A SETAR (k, p, d) is given by

$$Y_t = \alpha_0^{(j)} + \sum_{i=1}^{p} \alpha_i^{(j)} Y_{t-1} + \varepsilon_t^{(j)}, \text{ if } r_{j-1} \leq Y_{t-d} < r_j, j = 1, 2, ...k \tag{50}$$

where k denotes the number of different regimes; d is the **delay parameter** and $r_1 ... r_{k-1}$ are the **threshold parameters** which satisfy $-\infty = r_0 < r_1 <, ..., < r_k = \infty$; ϵ_t^j is white noise with ϵ_t^j independent of ϵ_t^l for $j \neq l$.

A more general case is when the model includes some moving average terms. A SETARMA (k, p, q, d) is given by

$$Y_t = \alpha_0^{(j)} + \sum_{i=1}^{p} \alpha_i^{(j)} Y_{t-1} + \sum_{i=1}^{q} \beta_i^{(j)} \varepsilon_{t-i} + \varepsilon_t \tag{51}$$

if

$$r_{j-1} \leq Y_{t-d} < r_j, \ \jmath = 1, 2, ..., k$$

These models give rise to phenomena such as limit cycles, catastrophe (jump phenomenon), asymmetries, corridor stability (state dependent response to shocks), thus capturing many forms of nonlinear interaction between variables.

Tsay (1989) provides an easy to implement procedure to test for the nonlinearities of SETAR type and to estimate $k, p_1, ..., p_k, d$ and $\{r_j\}s$. Applications to macroeconomic data can be found in the papers collected in Semmler (1994).[7]

[7] Potter (1993) found evidence for threshold nonlinearities in post-war U.S. GNP data and estimated a two regime SETAR model. Using nonlinear impulse response functions he found that asymmetries exist in GNP dynamics. Tiaoo & Tsay (1991) used a more refined regime categorization which distinguishes between upswing and downswing phases and estimated a four-regime TAR model for the same data. In both models the behavior of real GNP is different in recession and expansion periods. They both capture the notion of different stability properties in different regimes and asymmetric movements in the form of rapid turns into recessions (Potter, (1994), Cao & Tsay, (1992)). Cao & Tsay (1992) applied the SETAR modeling procedure developed by Tsay (1989) to daily stock return

An **Open-loop TAR (TARSO) model** (Tong, 1990) is characterized by the fact that the regimes are generated by a different variable X_{t-d},

$$Y_t = \alpha_0^{(j)} + \sum_{i=1}^{p} \alpha_i^{(j)} Y_{t-1} + \sum_{i=1}^{p} \beta_i^{(j)} X_{t-i} + \varepsilon_t^{(j)} \tag{52}$$

if

$$r_{j-1} \leq X_{t-d} < r_j,$$

Generalizations of this model include an $n-$equation TARSO model, called **TARS** and any mixture of SETAR and TARSO equations.

A further variant of this framework is elaborated by Luukonen et al. (1988a, 1988b), Granger and Teräsvirta (1993), Teräsvirta (1994). It is based upon the idea that, in contrast to SETAR models, although there is amplitude-dependence the transitions between regimes may take place smoothly.

A single equation STR-model can be written as

$$Y_t = \beta^{'} x_t + (\theta^{'} x_t) F(z_t) + \varepsilon_t \tag{53}$$

$$x_t = (1, Y_{t-1}, ...Y_{t-p};\ x_{1t}, ...x_{kt})^{'}$$

where z_t is any variable which is thought to be governing the transition between regimes and $F(z_t)$ some continuous function. One widely used form of STR model is **logistic STR(LSTR)model** characterized by logistic function F, i.e.,

$$F(z_t - c) = [1 + \exp(-\gamma(z_t - c))]^{-1}, \quad \gamma > 0 \tag{54}$$

$F(-\infty) = 0,\ F(+\infty) = 1,\ F(0) = 1/2.$

Another subcategory is called **exponential STR (ETSTR)model** where function F is exponential, i.e.,

$$F(z_t - c) = 1 - \exp[-\gamma(z_t - c)^2], \quad \gamma > 0 \tag{55}$$

$F(\pm\infty) = 1, F(0) = 0.$

If the transitions are generated by deviations of the transition variable from its linear (trend) path rather than from a fixed value c the model takes the form of **STR-DEVIATION (STR-D) model**. One possibility is to use lagged fitted residuals from the linear part of the equation (141) as the transition variable, i.e.,

$$z_t = Y_{t-d} - \hat{\beta}^{'} x_{t-d} \tag{56}$$

volatilities and found that SETAR models fare better in forecast performance as compared to GARCH and EGARCH models, especially for large stock returns.

then, depending upon the form of the transition function one could have a logistic or exponential **STR-D** model.[8] [10]

6.3 Estimation of STR Models

The empirical methodology to apply an STR model should be composed of the following steps:

1. Specify a linear model as a starting point: Use model selection criteria and residual diagnostic tests to specify an appropriate vector autoregression (VAR) model.

2. Test linearity against STR using the linear model specified in the first step as the null model. If the linearity hypothesis is rejected, determine the transition variable (or a linear combination of them) from the data.

    - A test with power against both LSTR and ESTR involves testing $H_0 : \emptyset_1 = \emptyset_2 = \emptyset_3 = 0$ in the following auxiliary regression:

    $$Y_t = \beta^{'} x_t + \emptyset_1^{'} x_t z_{td} + \emptyset_2^{'} x_t z_{td}^2 + \emptyset_3^{'} x_t z_{td}^3 + \eta_t$$

    The above regression can also be used to select the transition variable z_t by conducting the test for different variables.

    - If linearity is rejected for several choices of z_t, the select the one with the smallest probability value as the transition variable.

3. Choose between the LSTR and ESTR models. Consider the following sequence of nested hypothesis:

[8]Granger and Teräsvirta (1993), Teräsvirta (1994), and Eitrheim and Teräsvita (1995) present a full-fledge testing, specification, estimation und evaluation procedure for STR models and show that TR family, contains TAR, TARSO, TARSC, and Exponential Autoregressive (EXPAR)[9] models as special cases. Teräsvirta and Anderson (1992), Granger and Teräsvirta & Anderson (1993) and Teräsvirta (1993) applied this procedure to various economic series.

[10]In particular, Teräsvirta & Anderson (1992) tested and modeled **STAR**-STAR stands for smooth transition autoregressive and is distinguished by being univariate. STAR models constitute most of the empirical literature which uses STR methodology-specifications for quarterly industrial production indices of OECD countries, for the period between 1960.1 and 1986.4. For the U.S. data an LSTAR model is estimated which performs better than a linear one and whose dynamic behavior may include chaos and long and short-cycles. They found that recessions are locally explosive with complex roots and have a period of 8.9 quarters whereas expansions are stable with complex roots and periods of 61 quarters. Hence, the asymmetric behavior which has been observed for U.S. GNP in previous studies using threshold autoregressive models is also detected for industrial production series.

$$
\begin{aligned}
H_{03} &: \emptyset_3 = 0 \\
H_{02} &: \emptyset_2 = 0 \mid \emptyset_3 = 0 \\
H_{01} &: \emptyset_1 = 0 \mid \emptyset_3 = \emptyset_2 = 0
\end{aligned}
$$

If the test of H_{02} has the smallest probability value choose ESTR family, otherwise choose the LSTR family.

4. Estimation of a specified STR model can be undertaken by nonlinear least squares.

5. Evaluation:

   - Check if the parameter estimates are reasonable,
   - Diagnostic checks on residuals,
   - Evaluate the dynamic properties of the model.

An application of the above methodology to test a nonlinear model using the STR approach is given in Semmler and Kockesen (1997) for a dynamic economic model and in Chiarella, Semmler and Kockesen (1996) for a model in finance. A GAUSS program for a multivariate STR estimation is also available.

7 ARCH/GARCH-Models

7.1 Motivation and Derivation

In the financial literature nonlinear models became popular since stylized facts revealed that financial time series data do not appear to be consistent with the assumption of normal distribution. The so called ARCH/GARCH-models consider time-dependent variances of a time series.

Suppose the financial time series y_t follows a normal distribution with mean $\mu = 0$ and standard deviation h_t, an ARCH(p) model is formalized as follows:

$$
\begin{aligned}
y_t &= \epsilon_t h_t^{0.5}, y_t \mid y_{t-1} \sim N(0, h_t), \\
h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + ... + \alpha_1 y_{t-p}^2, a_1, ...a_p > 0.
\end{aligned}
$$

Note, that the squared y's are a consequence of the identity of the conditional expectation of y_t^2 and the variance of y in t. Practical investigations of the ARCH(p) model provide-evidence that p has to be chosen quiet large in order to obtain good approximation results.

This overparameterization may be avoided by employing an extended ARCH model, amounting to the following GARCH(p, q) model:

$$\begin{aligned} y_t &= \epsilon_t h_t^{0.5},\ y_t \mid \psi_{t-1} \approx N(0, h_t), \\ h_t &= \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{i=1}^{q} \beta_i y_{t-i}^2, \\ p, q &> 0 \text{ and } \alpha_1, \beta_j > 0, i = 1, ...p, j = 1, ..., q. \end{aligned}$$

Note, that the variance of y_t is only finite in case of $\sum \alpha_i + \sum \beta_j < 1$.

7.2 Estimation and Tests

Estimation of the structural parameter set $\theta = (\alpha, beta)$ is usually performed by the Maximum likelihood technique. For normally distributed y_t, one obtains the likelihood function

$$\mathcal{L}_t(\theta; y_t) = \frac{1}{\sqrt{h_t}\sqrt{2\pi}} \exp^{-\frac{1}{2h_t}y^2}.$$

Numerically, the likelihood function can be solved using the gradient $\frac{\partial \mathcal{L}}{\partial \theta}$.

In practical applications the optimal lag lengths p and q have to be chosen. This procedure may be guided by a test whether residuals exhibit further nonlinear dependencies using the test of Brock, Dechert and Scheinkmann.

Computer programs are available in GAUSS.

8 Neural Networks

8.1 Motivation und Derivation

Addressing the problem of nonlinearities, most widely employed nonlinear time series models are now nonparametric models like kernel regression as well as general approximation methods applying neural networks. The latter is a well known representative model in the econometric literature.

Assuming a financial time series generated by the conditional distribution $f(y_t \mid y_{t-1}, ...y_{t-m})$ can be represented by the stochastic relationship

$$y_t = g(y_{t-1}, ...y_{t-m}) + \epsilon_t, E(\epsilon_t) = 0, Var(\epsilon_t) < \infty,$$

where g is a nonlinear, but differentiable function neural networks are an appropriate nonlinear econometric method to estimate g, due to the universal approximation capabilities of the in θ parameterized function, the so called multi-layer perceptron:

$$\hat{y}_t = \hat{g}(\theta, y_{t-1}, ..., y_{t-m}) = \theta_0^1 + \sum_{h=1}^{H} \theta_h^1 \Psi(\theta^2, x_t)$$

$$\Psi(\theta^2, x_t) = (1 + exp(-\theta_{h0}^2 - \sum_{i=1}^{n} \theta_{hi}^2 x_{it})),$$

where Ψ is a logistic, i.e. sigmoid function.

8.2 Estimation and Tests

Employing the method of Backpropagation means to update an initial guess of the parameter set θ_0[11] iteratively using the gradient descent method,

$$\theta_i = \theta_{i-1} - \eta \frac{\partial MSE(y, \hat{y})}{\partial \theta}, i = 1, 2....,$$

with a small learning parameter η, performing parameter changes with maximal loss in the error function. Getting stuck in a local minimum in the error surface, several methods such as for example simulated annealing techniques are suggested to obtain convergence to the global optimum.[12]

An important task for the model builder is to choose the right complexity H of the neural network to avoid biased estimators (under-parameterization) as well as to prevent from having too much variance in the estimated data (over-parameterization). A straight-forward procedure would be the following:

1. fit the data $y_t, t = 1, ..., T - k$ for $H = 1, 2, ...$;

2. chose H^* so as to minimize $MSE_H(y, \hat{y})$;

3. fit the data $y_t, t = 1, ..., T$ for h^*.

More sophysticated techniques for neural network model selection are reviewed in Hertz, Krogh and Palmer (1993).

Computer programs are available in GAUSS and RATS.

[11] How to get an adequate initial guess of the parameters is discussed in Hertz, Krogh and Palmer (1993). In practice small random numbers are appropriate to avoid the tails of the logistic function.

[12] This and further topics of how to guide the learning process are reviewed in Hertz, Krogh and Palmer (1993).

Part IV: Estimation of Intertemporal Models

Next, we will sketch estimated strategies for intertemporal models where the dynamic model to be estimated is derived from the first order conditions of a dynamic optimization problem. We discuss discrete time and continuous time examples.

9 Estimating a Stochastic Dynamic Using MLE

9.1 The Model

A discrete time variant of a nonlinear intertemporal model[13] usually reads as follows:

$$V(x_0, u_0) = E_{x_0} \sum_t \beta^t F(x_t, u_t)$$

$$\text{s.t. } x_{t+1} = x_t + f(x_t, u_t) + \epsilon_{t+1} \tag{57}$$

where x_t is a vector of state variables, u_t a vector of control variables, β the discount rate, ϵ_{t+1} a vector of random shocks to the economy, $E_{x_0}, u_0 z_0$ is expectation conditioned on the information at time t and $F(x_t, u_t)$ the return function. This is the general form of a real business cycle (RBC) model. For details of the following model, see Semmler and Gong (1996).

Written in feed back form, in general, the control variables are nonlinear functions of the state variables.

$$u_t = G(x_t) \tag{58}$$

In special cases the nonlinear map (58) can explicitly be computed.[14] Using a Lagrangian approach a linearized version of an intertemporal model - linearized in the state variables, Lagrangian multipliers and control variables - is proposed by Chow (1993), see also Chow (1997). In general it can be written as

$$x_{t+1} = Ax_t + Cu_t + b + \epsilon_{t+1} \tag{59}$$

$$\lambda_{t+1} = Hx_t + h \tag{60}$$

$$u_{t+1} = Gx_t + g + e_t \tag{61}$$

[13] Model variants of the subsequent type are extensively treated in Cooley (1995).

[14] Semmler and Sieveking (1996), see also Taylor and Uhlig (1990) for a comparison of different techniques to compute the control variables in feedback form and to solve those models.

where

$$G = -(K_2 + \beta C' HC)^{-1}(K_{21} + \beta C' HA) \tag{62}$$

Furthermore,

$$g = -(K_2 + \beta C' HC)^{-1}[k_2 + \beta C' (Hb + h)] \tag{63}$$

where

$$h = (K_{12} + \beta A' HC)g + k_1 + \beta A' (Hb + h) \tag{64}$$

and

$$H = K_1 + K_{12}G + \beta A' H(A + CG) \tag{65}$$

λ represents the vector of Lagrangian multipliers. The other matrices and vectors are to be explained below. The parameters of the model (57) are embedded in the matrix G and the vector g.

Next, let us specify the above general intertemporal model as an RBC model[15] whereby the return function takes the form of a utility function including two control variables, consumption, u_1 and labor input, u_2

$$F(u_{1,t}u_{2,t}) = log\ u_{1,t} + \theta \log(1 - u_{2,t}) \tag{66}$$

The state space is two dimensional representing capital accumulation, x_2 and a stochastic process denoting a sequence of productivity shocks x_1. Employing a Cobb-Douglas production function we obtain

$$q_t = (x_{2,t})^{1-\alpha}(A_t u_{2,t})^{\alpha}$$

The dynamics of the two state variables are

$$x_{1,t} = \gamma + x_{1,t-1} + \epsilon \tag{67}$$

$$x_{2,t} = (1 - \delta)x_{2,t-1} + x_{2,t-1}^{1-\alpha} e^{\alpha x_{1,t-1}} u_{2,t-1}^{\alpha} - u_{1,t-1} \tag{68}$$

where in equ. (67) $x_{1,t} = log\ a_t$ representing a random walk with a drift, γ and ϵ, a random shock to the technology. Equ. (68) represents the evolution of the capital stock, $x_{2,t}$ and δ, being the depreciation rate. The structural parameters involved in this RBC model are

$$\varphi = (\alpha, \beta, \theta, \delta, \gamma).$$

[15] For details, see Chow (1993, 1997)

which are embedded in the matrices and vectors of Eqs. (62) - (65) which, in the case of an RBC model can be defined as

$$
\begin{aligned}
A &= \begin{bmatrix} 1 & 0 \\ \alpha & \bar{q}(1-\delta+(1-\alpha)\bar{q}/\bar{x}_2 \end{bmatrix} \text{ and } C = \begin{bmatrix} 0 & 0 \\ -1 & \alpha\bar{q}/\bar{u}_2 \end{bmatrix}; \\
b &= (\gamma, -\alpha\bar{q}\bar{x}_1)', \ \epsilon = (\epsilon_t, 0)', \ K_1 = 0, \ K_{12} = 0, \ k_1 = 0 \ K_{21} = 0; \\
k_2 &= (2\bar{u}_1^{-1}, \ -\theta[(1-\bar{u}_2)^{-1} - (1-\bar{u}_2)^{-2}\bar{u}_2])', \\
\text{and } K_2 &= \begin{bmatrix} -\bar{u}_1^{-2} & 0 \\ 0 & \theta(1-\bar{u}_2)^{-2} \end{bmatrix}
\end{aligned}
$$

Given the parameter set φ the linearized functions for the control variables, corresponding to equ. (58), read

$$
\begin{bmatrix} 1 & 0 & 0 \\ -G_{11} & 1 & 0 \\ -G_{21} & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1,t} \\ u_{1,t} \\ u_{2,t} \end{bmatrix} = \begin{bmatrix} \gamma \\ g_1 \\ g_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ u_{1,t-1} \\ u_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ G_{12} \\ G_{22} \end{bmatrix} x_{2,t} + \begin{bmatrix} \epsilon_t \\ e_{1,t} \\ e_{2,t} \end{bmatrix}
$$

9.2 Estimating the model

The above model can be written as standard simultaneous equation system

$$B_{y_t} + \Gamma x_t = e_t \tag{69}$$

where $y = (x_{1,2}, u_{1,t}, u_{2,t})', x_t = (x_{1,t-1}u_{1,t-1}, u_{2,t-1}, x_{2,t}1)'$ and

$$
B = \begin{bmatrix} 1 & 0 & 0 \\ -G_{11} & 1 & 0 \\ -G_{21} & 0 & 1 \end{bmatrix}, \Gamma = \begin{bmatrix} -1 & 0 & 0 & 0 & -\gamma \\ 0 & 0 & 0 & -G_{12} & -g_1 \\ 0 & 0 & 0 & -G_{22} & -g_2 \end{bmatrix}, e_t = \begin{bmatrix} \epsilon_t \\ e_{1,t} \\ e_{2,1t} \end{bmatrix}
$$

For n observations we obtain

$$BY' + \Gamma X' = E' \tag{70}$$

where Y' is $3 \times n, X'$ is $5 \times n$ and E' is $3 \times n$. Following Chow (1983: 170-171) and Chow (1993:13) we can introduce a concentrated log-likelihood function to iteratively compute the deep parameters φ, starting with $\varphi_0(\alpha_0, \beta_0, \theta_0, \delta_0, \gamma_0)$. It has the form

$$logL = const + nlog \mid b \mid -(n/2)log \mid (1/n)(BY' + X\Gamma')$$

whereby the maximum likelihood estimate is

$$\hat{\sum} = (1/n)(BY' + \Gamma X')(YB' + X\Gamma') \tag{71}$$

To compute the deep parameters iteratively it requires three steps

- start with initial values φ_0 and compute G, g, H and h of equs. (62)-(65),
- evaluate the log likelihood function (71) for φ_0,
- apply an optimization algorithm to change the φ_t in the direction of maximizing log L.

The recursive procedure which employs the simulated annealing as optimization algorithm is available in GAUSS. The results of the estimation for U.S. macro economic time series data 1952-1990are reported in Semmler and Gong (1996, 1997). Asset market restrictions can also be considered in intertemporal models of the above type, see Lettau, Gong and Semmler.

10 Estimating an Endogenous Growth Model Using GMM

Next we present a continuous time variant of an intertemporal model and sketch an estimation strategy.

10.1 The model

Consider an endogenous growth model in continuous time in the context of a closed economy[16]The economy is composed of three sectors: the household sector is assumed to optimize whereas the government sector does not optimize but follows certain budgetary rules.

The household is supposed to maximize the discounted stream of utilities arising from consumption subject to its per capita budget constraint, i.e.

$$\max_{C(t)} \int_0^\infty e^{-(p-n)t} L_0 u(C(t))dt,$$

subject to

$$\begin{aligned} &C(t) + \dot{K}(t) + (\delta_1 + n)K(t) + \dot{B}(t) + nB(t) \\ &= (w(t) + r_1(t)K(t) + r_2(t)B(t))(1-\tau) + T_p(t). \end{aligned}$$

[16]The analytical treatment of the subsequent model is given in Greiner and Semmler (1996a). The model is estimated in Greiner, Semmler and Gong (1997).

where $C(t)$ is consumption at time T. The assets accumulated by the household are physical capital $K(t)$, which depreciates at the rate δ_1 and government bonds or public debt $B(t)$. $T_p(t)$ is the lump-sum transfer payment to the household which the household takes as given in solving its optimization problem. The term τ is the income tax rate and $w(t), r_1(t)$, and $r_2(t)$ denote the wage rate, the return to physical capital, and the return to government bonds respectively. The no-arbitrage condition requires the after tax equalization of the two rates of returns (see appendix of Greiner, Semmler and Gong (2000)).

Moreover, ρ is the constant rate of time preference. The labor supply which equals L_0 at $t = 0$ is assumed to grow at the constant rate n.

As to the utility function we use the CRRA function

$$(C(t)^{1-\sigma} - 1)/(1 - \sigma),$$

where σ is the coefficient of relative risk aversion which is a constant. For $\sigma = 1$ the utility function can be replaced by the logarithmic function in $C(t)$.

The productive sector is assume to be represented by a firm which behaves competitively exhibiting a per capita production function of the form,

$$f(K, G) = K^{\beta}(\overline{G}/L)^{\alpha}.$$

where $\overline{G}$ is the aggregate stock of public capital which is subject to congestion. As to congestion we adopt the modeling proposed in Glomm and Ravikumar (1994) and assume that the per capita stock o public capital $G = \overline{G}/L$ affects per capita output. β and $\alpha, \beta, \alpha \in (0, 1)$ denotes the share of private and public per capita capital in the production function respectively. Since K denotes per capita capital the wage rate and the return to private capital are determined as $w = (1 - \beta)K^{\beta}G^{\alpha}$ and $r_1 = \beta K^{\beta-1}G^{\alpha}$. The budget constraint of the government in per capita terms is given by

$$\dot{B} = r_2 B + C_p + T_p + \dot{G} - T - nB.$$

where $r_2 B$ is the debt service, C_p stands for public consumption, T_p for transfers, $\dot{G}$ for public investment and T for the tax income, given by $T = \tau(w + r_1 K + r_2 B)$.

In addition as to the sustainability of public debt we posit that the government is not allowed to play a Ponzi game. We thus state that the usual transversality condition

$$\lim_{t \to \infty} B(t)e^{-\int_0^t (r_2(s) - n)ds} = 0$$

must hold. In Greiner and Semmler (1996) we show that the above transversality condition will hold when certain parameter constellations are given.[17]

[17] In Semmler and Sieveking (1996) the problem of sustainability of debt is analytically studied for a related model and in Greiner and Semmler (1997) the above sustainability condition is estimated for German time series data. Section 13 discusses details of such a test.

The dynamic behavior of our economy can be analyzed after defining the ratios $c = C/K,\ b = B/K$, and $x = G/K$. Differentiating c, b, and x with respect to time we get a new dynamic system which completely describes our model around a BGP. The dynamic system, where the growth rates of c, b, and x are derived from the first order condition of the above maximization problem by employing the Hamiltonian approach, is given by the following equations (76)-(78):

$$\begin{aligned}
\frac{\dot{c}}{c} &= -\frac{\rho+\delta_1}{\sigma} + \frac{(1-\tau)\beta G^\alpha}{\sigma K^{1-\beta}} + \frac{C}{K} + (\delta_1 + n) - \frac{G^\alpha}{K^{1-\beta}} \\
&\quad + \tau(\varphi_2 + \varphi_3(1-\varphi_0)) \cdot \Big(\frac{G^\alpha}{K^{1-\beta}} + \frac{B}{K}\Big(\beta\frac{G^\alpha}{K^{1-\beta}} - \frac{\delta_1}{1-\tau}\Big)\Big) \\
\frac{\dot{b}}{b} &= (\varphi_0 - 1)(1-\varphi_3)\tau\Big(\beta\frac{G^\alpha}{K^{1-\beta}} - \frac{\delta_1}{1-\tau} + \frac{K^\beta G^\alpha}{B}\Big) \\
&\quad + (1-\varphi_4)\Big(\beta\frac{G^\alpha}{K^{1-\beta}} - \frac{\delta_1}{1-\tau} + \frac{C}{K} + \delta_1 - \frac{G^\alpha}{K^{1-\beta}} \\
&\quad + \tau(\varphi_2 + \varphi_3)(1-\varphi_0)\Big(\frac{G^\alpha}{K^{1-\beta}} + \frac{B}{K}\Big(\beta\frac{G^\alpha}{K^{1-\beta}} - \frac{\delta_1}{1-\tau}\Big)\Big), \\
\frac{\dot{x}}{x} &= \varphi_3(1-\varphi_0)\tau\Big(\frac{K^\beta}{G^{1-\alpha}} + \beta\frac{G^\alpha}{K^{1-\beta}}\frac{B}{G} - \frac{B}{G}\frac{\delta_1}{1-\tau}\Big) - \delta_2 + \frac{C}{K} + \delta_1 \\
&\quad - \frac{G^\alpha}{K^{1-\beta}} + \tau(\varphi_2 + \varphi_3(1-\varphi_0))\Big(\frac{G^\alpha}{K^{1-\beta}} + \frac{B}{K}\Big(\beta\frac{G^\alpha}{K^{1-\beta}} - \frac{\delta_1}{1-\tau}\Big)\Big)
\end{aligned}$$

Here we have assumed that public consumption and transfer payments to the household constitutes a certain part of the tax income, i.e. $C_p = \varphi_2 T$ and $T_p = \varphi_1 T,\ \varphi_1, \varphi_2 < 1$. Moreover, we define per capita government expenditure for public investment as $\dot{G} = \varphi_3 \cdot (1-\varphi_0)T - (\delta_2 + n)G$, with $\varphi_3 \geq 0$ and δ_2 the depreciation rate of public capital. The fraction φ_0 depends on the policy.

For $\beta = 1 - \alpha$, the above is an autonomous system of differential equations in the variables c, b, and x. The local dynamics of the model can then be analyzed by computing analytically or numerically the eigenvalues of the Jacobian matrix. The above system may also exhibit multiple equilibria (for details, see Greiner, Semmler and Gong, 1997).

10.2 Estimating the Model

We employ time series data on consumption, public debt and public capital stock to estimate the above model for the U.S. and German economies from 1952 to 1990. All variables are defined relative to the private capital stock. We first need to describe our estimation strategy. Then we turn to the description of the actual estimation.

We employ a GM estimation and the Euler Scheme for the discretization of the continuous time model. The GMM estimation starts with a set of orthogonal conditions, representing the population moments established by a theoretical model:

$$E[g(y_t, \psi)] = 0 \tag{72}$$

where y_t is a $p \times 1$ vector of observed variables at date $t; \psi$ is a $q \times 1$ vector of unknown parameters to be estimated and $g(\cdot)$ is a $r \times 1$ vector mapping from R^{p+q}. Let T denote the sample size. The sample moments of $g(\cdot)$ can be written as

$$g_t(\psi) = \frac{1}{T}\sum_{t=1}^{T} g(y_t, \psi). \tag{73}$$

the idea of GMM estimator is to choose an estimated ψ that matches the sample moments $g_t(\psi)$ and the population moments given by (73) as closely as possible. To achieve this, one needs to define a distance function by which that closeness can be judged. Hansen (1982) suggested a distance function:

$$J_t(\psi) = [g_T(\psi)' W_T [g_T(\psi)], \tag{74}$$

where W_T, called the weighting matrix, is $r \times r$, symmetric and positive definite. Thus, the GMM estimator is the value of ψ, denoted as $\hat{\psi}$, that minimizes (75). From the results established in Hansen (1982), consistent estimator of the variance-covariance matrix of $\hat{\psi}$ is given by

$$Var(\hat{\psi}) = \frac{1}{T}(D_T)^{-1} W^{-1} T (D_T')^{-1}, \tag{75}$$

where $D_T = \partial g_T(\hat{\psi})/\partial \psi'$.

There is a great flexibility in the choice of W_T for constructing a consistent and asymptotically normal GMM estimator. We can adopt the method by Newey and West (1987), where it is suggested that

$$W_t^{-1} = \hat{\Omega}_0 + \sum_{j=1}^{m} w(j,m)(\hat{\Omega}_j + \hat{\Omega}_j'), \tag{76}$$

with $w(j,m) \equiv 1 - j/(1+m)$, $\hat{\Omega}_j \equiv (1/T)\sum_{t=j+1}^{T} g(y_t, \hat{\psi}^*) g(y_{t-j}, \hat{\psi}^*)$ and m to be a suitable of ψ. Thus two-step estimation is suggested as in Hansen and Singleton (1982). First, one chooses an sub-optimal weighting matrix to minimize (74) and hence obtains a consistent estimator $\hat{\psi}^*$. One then uses the consistent estimator obtained in the first step to calculate the optimum W_T through which (74) is re-minimized.

To define the set of orthogonal conditions (73) for our GMM estimation we can employ our above dynamic system. Then we have three equations defining our set of orthogonal conditions as follows:

$$E[\tilde{c} - f_1(c(\psi), x(\psi), b(\psi))] = 0 \tag{77}$$

$$E[\tilde{b} - f_2(c(\psi), x(\psi), b(\psi))] = 0 \tag{78}$$

$$E[\tilde{x} - f_3(c(\psi), x(\psi), b(\psi))] = 0 \tag{79}$$

where $f_i (i = 1, 2, 3)$ depends on the the parameter set $\psi = (\rho, \sigma, \alpha)$. the terms $\tilde{c}, \tilde{b}, \tilde{x}$ represent the deviation of the actual growth rates from their trend values at time period t.[18] The sample moments ($g(\cdot)$ as defined in (73) can be computed using (77)-(79).Our estimation problem with three equations and three parameters $\psi = (\rho, \sigma, \alpha)$ amounts to numerically finding such a parameter set ψ such that the distance as expressed in (72) is minimized.

A computer algorithm, written in "Gauss" is designed to solve the optimization problem (72) with the simulated annealing in the above mentioned two steps. The estimations are undertaken for U.S. and German time series data for 1952-1990. For details of the results, see Greiner, Semmler and Gong (1997).

11 Estimating Transversality Conditions of Intertemporal Models: Bubble Tests

Recently some effort has been spent to empirically test whether a transversality condition, for example as stated in sect. 12.1 can hold. this is equivalent to testing for bubble in time series data. We report here an estimator that originates in Flood and Garber (1980) which has been extended by Hamilton and Flavin (1986) and recently employed in Greiner and Semmler (1997).[19]

In the above model of sect. 12.1 the evolution of the state variable, appearing in the transversality condition, can be written in discrete time form as

$$B_t = (1 + r)B_{t-1} - S_t \tag{80}$$

where S_t is a flow variable (government surplus) and B as stock (government debt). In models for an open economy B can be viewed as external debt and S the trade surplus.

[18] In order to devoid detrending through procedures like the Hodrick-Prescott (HP) filter which may bias the estimation results we here use a formulation in terms of growth rates as suggested by our model (78)-(80). We, however, also undertook the estimation with HP-detrended data. The results turned out to be less reasonable than the below reported results using growth rates.

[19] Note, that in theory we can compute the transversality condition at the steady state for $t \to \infty$. In the subsequent section we want to estimate the transversality or non-explosiveness condition for a finite number of observations.

By recursive substitution forwards, equation (81) becomes

$$B_t = \sum_{i=t+1}^{N} \frac{S_i}{(1+r)^{i-t}} + \frac{(1+r)^t B_N}{(1+r)^N} \tag{81}$$

For a more proper treatment of forward solutions in expectations models, see Gourieroux and Montfort (1997, ch.12). When wealth holders expectations are that the borrowing behavior of the government or country is subject to the present value borrowing constraint

$$B_t = E_t \sum_{i=t+1}^{\infty} \frac{S_i}{(1+r)^{i-t}} \tag{82}$$

this is equivalent to requiring that the real supply of debt held by the public is expected to grow no faster than the interest rate

$$E_t \lim_{N\to\infty} \frac{B_N}{(1+r)^N} = 0 \tag{83}$$

If the government debt is constrained not to exceed constant, A_0, on the right hand side of (2), we then have from (82)

$$B_t = E_t \sum_{i=t+1}^{\infty} \frac{S_i}{(1+r)^{i-t}} + A_0(1+r)^t \tag{84}$$

and the empirical question of non-sustainability amounts to the problem whether the term $A_0(1+r)^t$ is positive an significantly different from zero. Following Flood and Garber (1980) and Hamilton and Flavin (1986) the following nonlinear least square test for sustainability can be employed

$$S_t = b_1 + b_2 S_{t-1} + b_3 S_{t-2} + b_4 S_{t-3} + \varepsilon_{2t} \tag{85}$$

$$\begin{aligned} B_t = b_5(1+r)^t + b_6 + \frac{(b_2 b + b_3 b^2 + b_4 b^3) S_t}{(1 - b_2 b - b_3 b^2 - b_4 b^3)} \\ + \frac{(b_3 b + b_4 b^2) S_{t-1}}{(1 - b_2 b - b_3 b^2 - b_4 b^3)} + \frac{(b_4 b) St-2}{(1 - b_2 b - b_3 b^2 - b_4 b^3)} + \varepsilon_{1t} \end{aligned} \tag{86}$$

Equations (85)-(86) should be jointly estimated by nonlinear least square. The term $b = 1/(1+r)$ is based on a constant interest rate.

Note, that in order to create non-sustainability of debt or for a bubble to exist the parameter b_5 should be positive and significantly different from zero. Estimation results of the sustainability of government debt for the U.S. and Germany are reported in Greiner and Semmer (1997). Similar bubble or non-sustainability tests have been undertaken for price level (Flood and Garber, 1990), foreign debt (Trehan and Walsh, 1991) and stock market dynamics.A computer program to estimate equations (85)-(86) is available in GAUSS.

References

[1] Baxter, M. and R. King (1995), "Measuring business cycles approximate band-pass filters for economic time series.", NBER working paper 50220.

[2] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroscedasticity", Journal of Econometrics 31, pp. 307-327.

[3] Campbell, J.Y., Lo, A.W. and A.C. MacKinlay (1996), "The Econometrics of Financial Markets", Princeton University Press, Princeton, N.J.

[4] Cao and Tsay (1992), "Nonlinear Time-Series Analysis of Stock Volatilities", Journal of Applied Econometrics 7, pp. 165-185.

[5] Chiarella, C., Semmler, W. and L. Kockesen (1996), "The Specification and Estimation of a Nonlinear Model of Real and Stock Market interaction", mimeo, New School for Social research, New York.

[6] Chow, G. (1983), "Econometrics", Mac Graw Hill, New York.

[7] Chow, G. (1993), "Statistical Estimation and Testing of a Real Business Cycle Model", Econometric Research Program, Research Memorandum, no. 365, Princeton: Princeton University.

[8] Chow, G. (1997), "Dynamic Economics", Oxford University Press, New York et al.

[9] Cooley, T. (1995), ed., "Frontiers of Business Cycle Research", Princeton Unversity Press, Princeton.

[10] Eitrheim and T. Teräsvirta (1995), "Testing the Adequacy of Smooth Transition Autoregressive Models", unpublished paper.

[11] Engle, E.F. (198), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", Econometrica 50, pp. 987-1007.

[12] Franke, R. and W. Semmler (1994), "The Financial-Real Interaction and Investment in the Business Cycle: Theory and Empirical Evidence", in: Nell, E. and G. Deleplace (eds.), Money in Motion, Macmillan.

[13] Franke, R. and W. Semmler (1993), "Detrending Programs", (available on the DOS-level.).

[14] Glomm, G. and B. Ravikumar (1994), "Public Investment in Infrastructure in a Simple Growth Model", Journal of Economic Dynamics and Control, vol. 18, pp. 1173-87.

[15] Gourieroux, C. and A. Montfort (1997), "Time Series and Dynamic Models", Cambridge University Press, Cambridge et al.

[16] Granger, C.W. and T. Teräsvirta (1993), "Modelling Nonlinear Economic Relationships", Chicago University Press, Chicago.

[17] Granger, C.W., Teräsvirta, T. and H.M. Anderson (1993), "Modelling Nonlinearity over the Business Cycle", New Research on Business Cycles, Indicators and Forecasting, J.H. Storch and M.W. Watson (eds.), Chicago University Press, Chicago.

[18] Greiner, A. and W. Semmler (1996a), "Endogenous Growth, Government Debt and Budgetary Regimes", paper presented at the 52nd Congress, International Institute of Public Finance, Tel Aviv, Israel, August 1996, mimeo, University of Augsburg, New School for Social Research.

[19] Greiner, A., Semmler, W. and G. Gong (1997), "Estimating an Endogenous Growth Model with Public Capital and Government Borrwowing", University of Bielefeld, Working paper340.

[20] Hamilton, J.D. (1994), "Time Series Analysis", Princeton University Press, Princeton, N.J.

[21] Hansen, G. (1993), "Quantitative Wirtschaftsforschung", Verlag Vahlen, München.

[22] Hansen, L.P. (1982), "Large Sample Properties of Generalized Methods of Moments Estimators", Econometrica 50, pp. 1029-1054.

[23] Hansen, L.P. and K.J. Singleton (1982), "Generalized Instrument Variables Estimation of Nonlinear Expectations Models", Econometrica 50, pp. 1268-1286.

[24] Hertz, J.A., Krogh, A.S. and R.G. Palmer (1995), "Introduction to the Theory of Neural Computation", Addison Wesley Publishing Company, Redwood City, C.A. et al.

[25] Hodrick, ? and ? Prescott (1980), "Post-War U.S. Business Cycles", Working paper, Carnegie Mellon University.

[26] Kloeden, P.E. and E. Platen (1995), "Numerical Solutions of Stochastic Differential Equations", Springer Verlag, Heidelberg et al.

[27] Kloeden, P.E., Platen, E. and H. Schurz (1991), "Numerical Solutions of SNDE Through Computer Experiments", Springer Verlag, Heidelberg et al.

[28] Lettau, M., Gong, G. and W. Semmler (1997), "Parameter Estimation and Moment Evaluation of a Stochastic Growth Model with Asset Market".

[29] Luukonen, R., Saikkonen, P. and T. Teräsvirta (1988a), "Testing Linearity Against Smooth Transition Autoregression Models", Biometrica 75, pp. 491-499.

[30] Luukonen, R., Saikkonen, P. and T. Teräsvirta (1988b), "Testing Linearity in Univariate Time Series Models", Scandinavian Journal of Statistics 15, pp. 161-175.

[31] Neftci, S. (1996), "An Introduction to the Mathematics of Financial Deravatives", Academic Press, New York.

[32] Nelson, C. and C. Plosser (1982), "Trends and Random Walks in a Macroeconomic Time Series", Journal of Monetary Economics 10, pp.139-162.

[33] Newey, W.K. and K.D. West (1987), "A Simple, Positive-Definite, Heteroscedasticity and Autocorrelation Consistent Co-variance Matrix", Econometrica, vol. 55, pp. 703-708.

[34] Ozaki, T. (1985), "Nonlinear Time Series Models and Dynamical Systems", Handbook of Statistics 5, E.J. Hannan,P.R. Krishnaih and M.M. Rao (eds.), Elsevier Science Publishers B.V., North-Holland.

[35] Ozaki, T. (1986), "Local Gaussian Modeling of Stochastic Dynamical Systems in the Analysis of Nonlinear Random Vibrations", Essays in Time Series and Allied Processes: Papers in Honour of E.-J. Hannon, J. Gani and M.B. Priestley (eds.), Applied Probability Trust Sheffield.

[36] Ozaki, T. (1987), "Statistical Identification of Nonlinear Dynamics in Macroeconomics Using Nonlinear Time Series Models".

[37] Ozaki, T. (1989),"Statistical Identification of Nonlinear Random Vibration Systems", Journal of Applied Mechanics 56, pp. 186-191.

[38] Ozaki, T. (1994), "The Local Lineariation Filter with Application to Nonlinear System Identification", Proceedings of the First U.S./Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach, H. Bozdogan (ed.), Kluwer Academic Publishers,Boston, pp. 217-240.

[39] Potter, S.M. (1993), "A Nonlinear Approach to U.S. GNP", mimeo, University of Wisconsin-Madison.

[40] Potter, S.M. (1994), "Asymmetric Economic Propagation Mechanisms", Business Cycles: Theory and Empirical Methods, W. Semmler (ed.), Kluwer Academic Publishers, Boston.

[41] Semmler, W. (1994), Business Cycles: Theory and Empirical Methods, Kluwer Academic Publishers, Boston.

[42] Semmler, W. and M. Sieveking (1996), "Computing Creditworthiness and Sustainable Debt", paper prepared for he International Conference on Computing in Economics and Finance, University of Geneva, June 1996, mimeo, Department of Mathematics, Unversity of Frankfurt.

[43] Stock, J. and M. Watson (1997), "Business Cycle Fluctuations in U.S. Macroeconomic Time Series", forthcoming, Handbook of Macroeconomics, ed. by J. Taylor and M. Woodford.

[44] Taylor, J. and H. Uhlig (1990), "Solving nonlinear stochastic Growth Models: A Comparison of alternative solution of models", Journal of Business and Economic Statistics 8, pp. 1-17.

[45] Teräsvirta, T. (1994), "Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models", Journal of the American Statistical Association 89.

[46] Teräsvirta, T. and H.M. Anderson (1992), "Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models", Journal of Applied Econometrics 7, pp. S119-S136

[47] Tong, H. (1990), "Non-Linear Time Series, A Dynamical System Approach", Oxford University Press, New York.

[48] Tsay (1989), "Testing and Modeling Threshold Autoregressive Processes", Journal of the American Statistical Association 84, pp. 231-240