1

Chapter 6

Nonlinear Time Series Modelling

STRUCTURE

1. Introduction.
2. Threshold Autoregressive (TAR) models.
3. Markov Switching Models.
4. Nonlinearity test.
5. Forecasting.
6. Impulse Response Functions.
7. Models with Stochastic Volatility
8. Smooth transition regression modeling

Introduction

We focus on univariate time series, x_t. We denote the observations by $\{x_t\}_{t=1}^{n}$, where n is the sample size. The time series x_t is said to be linear if it can be written as

$$x_t = c + \sum_{i=0}^{\infty} \psi_i a_{t-1}$$

where c is a constant, $\psi_0 = 1$, ψ_i are real numbers and $\{a_t\}$ is a white noise process, $a_t \sim N(0, \sigma_a^2)$. If $\sigma_a^2 \sum_{i=1}^{\infty} \psi_i^2 < \infty$. Then x_t is weakly stationary.

Any process that does not satisfy this condition is said to be nonlinear.

There is a substantial interest in studying test statistics that can discriminate linear series from the nonlinear ones. Both parametric and nonparametric tests are now available.

The application of nonlinear models is less studied in the literature. Contributions of nonlinear models in forecasting are discussed only in recent years. Several factors contribute to this lack of progress:

1) There exist no commercially available statistical packages that can handle various nonlinear models.
2) No generally agreed upon methods are available to judge the real contributions of nonlinear models over linear ones. This is particularly so in forecasting.

1. THRESHOLD AUTOREGRESSIVE (TAR) MODELS

This model is motivated by several nonlinear characteristics commonly observed in practice, such as asymmetry in declining and rising patterns of a process.

It uses piecewise linear models to obtain a better approximation of the conditional mean equation.

However, unlike the traditional piecewise linear model that allows for model change to occur in the “time” space, the TAR model uses threshold space to improve linear approximation.

A time series x_t is said to follow a k-regime self-exciting TAR (SETAR) model with threshold variable x_{t-d} if it satisfies

$$\left(1-\phi_1^{(j)}L-\cdots-\phi_p^{(j)}L^p\right)x_t = c^{(j)}+a_t^{(j)},\ \ if\ r_{j-1}\leq x_{t-d}<r_j$$

where k and d are positive integers, $r_i{}'s$ are real numbers such that $-\infty = r_0 < r_1 < \cdots < r_k = \infty$.

2. Markov Switching Models

Using ideas similar to TAR models, but emphasizing aperiodic transition between various states of an economy.

A time series x_t follows a MSA model if it satisfies

$$x_t = \begin{cases} c_1 + \sum_{i=1}^{p} \phi_{1,i}\, x_{t-i} + a_{1t} \;\; if \;\; s_t = 1 \\ c_2 + \sum_{i=1}^{p} \phi_{2,i}\, x_{t-i} + a_{2t} \;\; if \;\; s_t = 2 \end{cases}$$

where s_t assumes values {1,2} and is a first-order Markov chain with transition probabilities

$$P(s_t = 2/s_{t-1} = 1) = w_1, \;\; P(s_t = 1/s_{t-1} = 2) = w_2$$

A small w_i means that the model tends to stay longer in state i.

- From the difinition, a MSA model uses a hidden Markov chain to govern the transition from one conditional mean function to another. This is different from that of a SETAR model, for which the transition is determined by a particular lagged variable. Consequently, a SETAR model uses a deterministic scheme to govern the model transition, whereas a MSA model uses a stochastic scheme.

- In practice, this implies that one is never certain about which state x_t belongs to in a MSA model.

- This difference has important practical implications in forecasting. For instance, forecasts of a MSA model are always linear combinations of forecasts produced by submodels of individual states.

NONLINEARITY TEST

Because nonlinearity may occur in so many ways, there exists no single test that dominates others in detecting nonlinearity

NON PARAMETRIC TESTS

Under the null hypothesis of linearity, the residuals of a properly specified linear model should be independent.

Q-STATISTIC OF SQUARED RESIDUALS

$$Q(m) = n(n+2)\sum_{i=1}^{m}\frac{\hat{\rho}_i^2(r_t^2)}{n-i}$$

where m is a properly chosen number of autocorrelations used int he test and r_t denotes the residual series. This test is particularly useful in detecting conditional heteroskedasticity of x_t. In particular, the Lagrange test statistic of Engle (1982) for ARCH models is closely related to the Q-statistics.

In essence, Engle considers the linear regression

$$r_t^2 = \beta_0 + \sum_{i=1}^{m} \beta_i \, r_{t-i}^2 + e_t$$

The null hypothesis (no conditional heteroskedasticity) is $H_0\colon \beta_1 = \cdots = \beta_m = 0$. Under the null, the F-ration follows an F-distribution with degrees of freedom m and $n - 2m - 1$. Asymptotically, this F-statistic has a chi-square distribution with m degrees of freedom

Parametric Test

THE RESET TEST

Ramsey (1969) proposes a specificationt test for linear least squares regression analysis. Consider the linear AR(p) model

$$x_t = X_{t-1}^{'}\phi + a_t$$

where $X_{t-1} = (1, x_{t-1}, \dots, x_{t-p})'$.

The first step of the RESET test is to obtain the least squares estimate $\hat{\phi}$, the residual $\hat{a}_t = x_t - \hat{x}_t$, and the sum of squared residuals $SSR_0 = \sum_{t=p+1}^{n} \hat{a}_t^2$.

In the second step, consider the linear regression

$$\hat{a}_t = X_{t-1}^{'}\alpha_1 + M_{t-1}{}'\alpha_2 + v_t$$

where $M_{t-1} = (\hat{x}_t^2, \dots, \hat{x}_t^{s+1})'$ for some $s \geq 1$, and compute the least squares residuals $\hat{v}_1 = \hat{a}_t - X_{t-1}^{'}\alpha_1 - M_{t-1}{}'\alpha_2$ and the sum of squared residuals $SSR_1 = \sum_{t=p+1}^{n} \hat{v}_t^2$. Because $\hat{x}_t^k$ for $k = 2, \dots, s+1$ tend to be highly correlated with X_{t-1} and among themselves, principal components of $\hat{x}_t^k$ that are not collinear with X_{t-1} are often used in this step.

In this case, the basic idea of RESET test is that if the linear AR(p) model is adequate then α_1 and α_2 should be zero. This can be tested by the usual F-statistic of equation given by

$$F = \frac{(SSR_0 - SSR_1)/g}{SSR_1/(n-p-g)}$$

with $g = s + p + 1$

which, under the linearity assumption, has an F-distribution with degrees of freedom g and $n - p - g$.

Kennan (1985) proposes a nonlinearity test that uses $\hat{x}_t^2$ only and modifies the second step to avoid multicollinearity between $\hat{x}_t^2$ and X_{t-1}. Specifically, the second step is divided into two steps. First, one removes linear dependence of $\hat{x}_t^2$ and X_{t-1} by fitting the regression

$$\hat{x}_t^2 = X_{t-1}^{'}\beta + u_t$$

And obtaining the residual $\hat{u}_t = \hat{x}_t^2 - X_{t-1}^{'}\hat{\beta}$.

In a second step, considet the linear regression

$$\hat{a}_t = \hat{u}_t\alpha + v_t$$

And obtain the sum of squared residuals $SSR_1 = \sum_{t=p+1}^{n}(\hat{a}_t - \hat{u}_t\hat{\alpha})^2$ to test the null hypothesis $\alpha = 0$.

TSAY'S TEST

To improve the power of of Kennan's and RESET tests, Tsay (1986) uses a different choice of the regressor M_{t-1}. Specifically, he suggests using $M_{t-1} = vech(X_{t-1}X_{t-1}^{'})$, where $vech$ denotes a half-stacking vector of the matrix $X_{t-1}X_{t-1}^{'}$ using elements on and below the diagonal only. The dimensión of M_{t-1} is then $p(p+1)/2$. In practice, the test is simply the usual F-statistic for testing $\alpha = 0$ in the linear squares regression

$$x_t = X_{t-1}^{'}\phi + M_{t-1}^{'}\alpha + e_t$$

Under the linearity assumption, the partial F-statistic follows a F-distribution with degrees of freedom g and n-p-g-1 where g=p(p+1)/2.

THRESHOLD TEST

When the alternative model under study is a SETAR model, one can derive specific test statistics to increase detecting power.

- One of the specific test is the likelihood ratio test. This test, however, encounters the difficulty of undefined parameters under the null hypothesis of linearity.

- Another specific test seeks to transform testing threshold nonlinearity into detecting model changes.

It is then interesting to discuss the differences between these two specific tests.

Let us consider the simple case that the alternative is a two-regime SETAR model with given threshold variable x_{t-d}. The null hypothesis is $H_0: x_t$ follows the linear AR(p) model

$$x_t = \phi_0 + \sum_{i=1}^{p} \phi_i x_{t-i} + a_t,$$

whereas the alternative hypothesis is $H_a: x_t$ follows the SETAR model

$$x_t = \begin{cases} \phi_0^{(1)} + \sum_{i=1}^{p} \phi_i^{(1)} x_{t-i} + a_{1t} & if \quad x_{t-d} < r_1 \\ \phi_0^{(2)} + \sum_{i=1}^{p} \phi_i^{(2)} x_{t-i} + a_{1t} & if \quad x_{t-d} \geq r_1 \end{cases}$$

where r_1 is the threshold.

Let $l_0(\hat{\phi}, \hat{\sigma}_a^2)$ be the log-likelihood evaluated at the maximum likelihood estimate. The likelihood function under the alternative is also easy to compute if the threshold r_1 is given. Let $l_1(r_1, \hat{\phi}_1, \hat{\sigma}_1^2, \hat{\phi}_2, \hat{\sigma}_2^2)$. The log –likelihood ratio is defined as

$$l(r_1) = l_1\left(r_1, \hat{\phi}_1, \hat{\sigma}_1^2, \hat{\phi}_2, \hat{\sigma}_2^2\right) - l_0(\hat{\phi}, \hat{\sigma}_a^2)$$

Is then a function of the threshold r_1, which is unknown. Furthermore, under the null hypothesis, there is no threshold and r_1 is not defined. For this reason, the asymptotic distribution of the likelihood ratio is rather different from that of the usual test; see Chan (1991) for critical values of the test.

A common approach is to use $l_{max} = sup_{v<r_1<u} l(r_1)$ as the test statistic, where u and v are pre-specified lower and upper bounds of the threshold; see Davis (1987) and Andrews and Ploberger (1994).

Simulation is often used to obtain empirical critical values of the test statistic l_{max} .

An average of $l(r_1)$ is also considered by Andrews and Ploberger as a test statistic.

Tsay (1989) makes use of arranged autoregression and recursive estimation to derive an alternative test for threshold nonlinearity.
He arranged autoregression seeks to transfer the SETAR model in the alternative hypothesis H_a into a model change problem with the threshold r_1 serving as the change point.

A SETAR model with a single threshold can be seen as to linear models depending on the position of variable x_t with respect to a threshold parameter. Let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n-d)}$ be the ordered statistics of $\{x_t\}_{t=1}^{n-d}$, the SETAR model can be written as

$$x_{(j)+d} = \beta_0 + \sum_{i=1}^{p} \beta_i x_{(j)+d-i} + a_{(j)+d}, \quad j = 1, \dots, n-d$$

where $\beta_i = \phi_i^{(1)}$ if $x_{(j)} < r_1$ and $\beta_i = \phi_i^{(2)}$ if $x_{(j)} \geq r_1$. Consequently, the threshold r_1 is a change point for the linear regression.

Note that <u>the arranged autoregression does not alter the dynamic dependence of x_t on x_{t-i} for $i = 1, \ldots, p$ because $x_{(j)+d}$ still depend on $x_{(j)+d-i}$ for $i = 1, \ldots, p$.</u>

What is done is to present the SETAR model in the threshold space instead of in the time space. That is, the equation with a smaller x_{t-d} appears before that with a larger x_{t-d}.

The threshold test of Tsay (1989) is obtained as follows:

1) Fit the equation using m observations, where m is a pre-specified positive integer, for example, 30. Denote the least squares estimates of β_i by $\hat{\beta}_{i,m}$

2) Compute the predictive residuals

$$\hat{a}_{(m+1)} = x_{(m+1)+d} - \beta_{0,m} - \sum_{i=1}^{p} \beta_{i,m} x_{(m+1)+d-i}$$

and its standard derror. Let $\hat{e}_{(m+1)+d}$ be standarized predictive residual.

3) Use the recursive least squares method to update the least squares estimates of $\hat{\beta}_{i,m+1}$ by incorporating the new data point.

4) Repeat steps 2 and 3 until all data points are processed.

5) Consider the linear regression of the standarized predictive residuals

$$\hat{e}_{(m+1)+d} = \alpha_0 + \sum_{i=1}^{p} \alpha_i \, x_{(m+1)+d-i} + v_t$$

and compute the usual F-statistic for testing $\alpha_i = 0$. Under the null hypothesis that x_t follows a linear AR(p) model, the F-ratio has a limiting F-distribution with degrees of freedom p+1 and n-d-m-1.

TAR test: under the null hypothesis there is no model change in the arranged autoregression so the standarized predictive residuals should be close to iid with mean zero and variance 1. In this case, they should have no correlation with the regressors $x_{(m+j)+d-i}$.

This F-test avoids the problem of nuisance parameters econunter in the likelihood ratio testl. It does not require knowing the threshold r_1. In fact, it does not depend on knowing that there is a single threshold in the model.

On the other hand, the F-test is not as powerful as the likelihood ratio test if the model is indeed a two-regime SETAR model with a known innovation distribution.

Forecasting

Unlike the linear model, in many cases there exist no closed-form formulae to compute forecasts of nonlinear models when the forecast horizon is greater than 1. Thus, we use parametric bootstrap to compute nonlinear forecasts.

Let T the forecast origin and l the forecast horizon ($l > 0$). The parametric bootstrap considered computes realizations $x_{T+1}, \ldots, X_{T+l}$ sequentially by

(a) Drawing a new innovation from the proper innovation distribution of the model and

(b) Computing x_{T+i} using the model, the data, and previous forecast $x_{T+1}, \ldots, x_{T+i-1}$. This result in a realization of x_{T+l}. This procedure is repeated M times to obtain M realizations of x_{t+l} denoted by $\left\{x_{t+l}^{(j)}\right\}_{j=1}^{M}$.

The point forecast of x_{t+l} is then the sample averages of the M realizations.

M can be equal to 3,000.

The realizations $\left\{x_{t+l}^{(j)}\right\}_{j=1}^{M}$ can also be used to obtain an empirical distribution of x_{t+l}.

IMPULSE RESPONSE FUNCTIONS

- Once one is satisfied with the nonlinear model there is the remaining question of describing how its dynamics differ from that of linear models fit to the same time series.
- The coefficients in the Wold Representation can be though of as producing the same answer to the following four questions:
  1) What is the response to a unit impulse today when all the future shocks are sent to zero.
  2) What is the response to a unit impulse today when all the future shocks are integrated out?
  3) What is the derivative of the predictor of the future?
  4) How does the forecast of the future change between today and yesterday, normalizing the change by the innovation to the time series today.

For nonlinear models there will be different answers to each question. Consider the simple threshold model

$$y_t = -1(y_{t-1} < 0) + 1(y_{t-1} \geq 0) + a_t$$

and the case of a positive unit impulse for the first two questions.

1) If $y_t \geq 0$ then the response is 0 for all the horizons, if $-1 \leq y_t < 0$ then response is 2 for all horizons since this permanently moves the time series into the upper regime given the assumption of no future shocks, if $y_t < -1$ then the response is 0 for all the horizons.
2) At horizon 1 the response is the same as the answer to question 1 but as the horizon increases we have the difference:
$$E[y_{t+h} \backslash Y_t = y_t + 1] - E[y_{t+h} \backslash Y_t = y_t]$$
which must converge to zero by the stationarity of the underlying time series.
3) The derivative is either 0 or not defined if $y_t = 0$.
4) Consider the case where $E_{t-1}[y_t] = 1$ and the realized value of $y_t = 1$. Then the initial shock is 0 but
$$E_t[y_{t+1}] - E_{t-1}[y_{t+1}] = 1 - 0.68 = 0.32$$
Or the case where $E_{t-1}[y_t] = 1$ and the realized value of $y_t = 5$. Then the initial shock is 4 but the $E_t[y_{t+1}] - E_{t-1}[y_{t+1}]$ is still equal to 0.32.

- Questions 1 and 3 are not particularly useful questions for ask for nonlinear series.
- This leaves a choice between the more traditional definition of the impulse response function defined by the answer to question 2 and the forecasting revision function defined by the answer to question 4.
- In order to choose between the two possibilities observe that both the initial condition and the magnitude and sign of the impulse is important in describing the dynamics of nonlinear models. This is problematic since we can chose values of the initial condition or shock that produce atypical responses.

- In order to define the initial contions one can use the history of the time series or random draws from its simulated distributions.
- In answering question 2 there is no direct way of defining the relevant set of perturbations away from the initial conditions using the properties of time series models
- Koop, Pesaran and Potter (1996) and Potter (1999) call the forecast revision function a generalized impulse response function and develop its properties.
- Multivariate extentions require to consider the fact that residuals in the different equations are contemporaneous. Atanasova (2000) and Tena and Tremayne (2009).
- Koop (1996) and Koop and Potter (1999) combine the analysis of impulse response functions under parameter uncertainty.

Models with Stochastic Volatility

REFERENCES

- Bollerslev, Tim. 1986. "Generalized Autoregressive Conditional Heteroscedasticity". Journal of Econometrics, 31, 307-27.
- Engle, R. 1982. "Autoregressive Condicional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", Econometrica, 50, 987-1007.
- Engle, R., D. Lilien, and R.Robins. 1987. « Estimating Time Variying Risk Premia in the Term Structure : the ARCH-M Model ». Econometrica, 55, 391-407.
- Holt, M. and S. Aradhyula. 1990. "Price Risk in Supply Equations: An Application of GARCH Time Series Models to the U.S. Broiler Market." Southern Economic Journal, 57, 230-42.

STRUCTURE

1) Introduction.
2) ARCH models.
3) GARCH models.
4) An example.
5) Models for Risk. ARCH-M.
6) Properties of GARCH(1,1) models.
7) Diagnosis and forecast.

1. INTRODUCTION

- Series do not have a constant mean or variance.
- Series with a constant variance are called homocedastic whereas series with a nonconstant variance are called heteroskedastic.
- So far, we have analysed how to capture the evolution of the mean with statistical models.
- However, more sofisticasted models require to propose a model for the mean and the varianze.

1. Introduction

- Some stilized facts of economic time series:
  1) Most of them show systematic growth while others show high persistence in different mean values for the different periods. We have seen different models to capture this behaviour.
  2) Economic series very often co-move with other series. We call this phenomenum: cointegration.
  3) The volatility of time series is not constant in time. We usually call this series heteroskedastic if the unconditional variance of these series is constant but the variance is very high during certain periods.

     To understand this is very important to distinguish between conditional and unconditional mean and variance.

1. Introduction

- Heteroskedastic series have different conditional and unconditional variances.
- The conditional variance can be forecasted one period ahead. One way to do it is to introduce in the model an exogenous variable that helps to forecast volatility.
- Consider this simple case:

$$y_{t+1} = x_t a_{t+1}$$

where y_t and x_t are the endogenous and exogeneous variables and a_t is a white noise process.

1. Introduction

- If $x_t = x_{t-1} = \ldots =$ constant, then y_t is a white noise process with constant variance.
- But, when x_t is not constant, the variance of y_t conditional to the observed value of x_t is

$$\mathrm{Var}(y_{t+1}) = x_t^2 \sigma^2$$

- Now, the conditional variance of y_t depends on x_t.

1. Introduction

- In practice, we will be interested in estimating models like:

$$\ln y_t = \alpha + \beta \ln x_{t-1} + a_t$$

where α and β can be estimated by OLS.

- A major difficulty in these types of models is that it assumes a specific cause for the change in variance. Sometimes this is not very clear. For example, was the oil shock or the monetary policy responsable of the output fluctuation after 1980?

2. ARCH MODELS

- Instead of using ad-hoc specification for x_t, Engle (1982) proposes a model that simultaneously captures the mean and variance. Assume I estimate the following stationary AR model

$$y_t = \alpha + \beta y_{t-1} + e_t$$

The conditional forecast error is

$$E(y_t - \alpha - \beta y_{t-1})^2 = E(e_t^2) = \sigma^2$$

The unconditional forecast is

$$E(y_t) = \frac{\alpha}{1-\beta}$$

2. ARCH MODELS

- The variance of the unconditional forecast error is:

$$E_t\left\{\left(y_{t+1}-\frac{\alpha}{1-\beta}\right)^2\right\}=$$

$$E_t\left[\left(e_{t+1}+\beta e_t+\beta^2 e_{t-1}+\beta^3 e_{t-2}+\ldots\right)^2\right]=\frac{\sigma^2}{(1-\beta^2)}$$

- The variance of the unconditional forecast error is always bigger than the variance of the conditional forecast error.

2. ARCH MODELS

- In a similar way, when the variance of a_t is not constant, we can propose a model for it

$$\hat{e}_t^2 = \alpha_0 + \alpha_1 \hat{e}_{t-1}^2 + \alpha_2 \hat{e}_{t-2}^2 + \ldots + \alpha_q \hat{e}_{t-q}^2 + v_t .$$

where v_t is a white noise process.

- This model is called autoregresive conditional heteroskedastic (ARCH)model.
- If the values α_1, α_2, …,α_q are zero, the conditional variance is constant and equal to α_0.

2. ARCH MODELS

- In practice, the mean and variance of y_t are simultaneously estimated. Besides, the element v_t is included in a multiplicative way.
- Engle (1982) proposes the following model

$$e_t = v_t \sqrt{\alpha_0 + \alpha_1 e_{t-1}^2}$$

with the assumptions

$$E(e_t, v_t) = 0; v_t \sim \mathbf{N(0,1);} \alpha_{\mathbf{0}} > 0; 0 < \alpha_1 < 1.$$

2. ARCH MODELS

- This model has the following properties

$$E(e_t) = E\left[\left(v_t \sqrt{\alpha_0 + \alpha_1 e_{t-1}^2}\right)^{1/2}\right] =$$

$$Ev_t E\left(\alpha_0 + \alpha_1 e_{t-1}^2\right)^{1/2} = 0.$$

- Besides, given that $E(v_t, v_{t-i})=0$, this means that

$$E(e_t, e_{t-i}) = 0 \qquad i \neq 0.$$

2. ARCH MODELS

- To obtain the unconditional variance of e_t

$$E(e_t^2) = Ev_t^2 E\left(\alpha_0 + \alpha_1 e_{t-1}^2\right) = \frac{\alpha_0}{1-\alpha_1}$$

- While the conditional variance is:

$$E(e_t^2 / e_{t-1}^2, e_{t-2}^2, \ldots) = \alpha_0 + \alpha_1 e_{t-1}^2$$

- That is, the conditional variance of e_t depends on the conditional variance of e_{t-1}. In other words, the variance follows an autoregressive process ARCH(1).
- The values of α_0 and α_1 are restricted to ensure that the variance is always positive and the process is stable.

ARCH MODELS

- Summary: in ARCH models, the unconditional mean of the error term is zero, moreover e_t is serially uncorrelated. However, they are not independent as they are related through their second moments.
- Formally, the model is

$$E_{t-1}y_t = \alpha + \beta y_{t-1}$$

$$\mathrm{Var}(y_t / y_{t-1}, y_{t-2}, \ldots) = E_{t-1}(y_t - a_0 - a_1 y_{t-1})^2 = E_{t-1}(e_t)^2 = \alpha_0 + \alpha_1 (e_{t-1})^2$$

ARCH MODELS

- The unconditional mean of y_t can be written as

$$y_t = \frac{\alpha}{1-\beta} + \sum_{i=0}^{\infty} \beta^i e_{t-i}.$$

- From this it is possible to obtain the unconditional variance of y_t

$$\mathrm{Var}(y_t) = \sum_{i=0}^{\infty} \beta^{2i} \mathrm{var}(e_{t-i})$$

- If the unconditional variance of y_t is constant,

$$\mathrm{Var}(e_t) = \mathrm{Var}(e_{t-1}) = \ldots = \frac{\alpha_0}{(1-\alpha_1)}$$

2. ARCH MODELS

- Then

$$\mathrm{Var}(y_t) = \left[\frac{\alpha_0}{(1-\alpha_1)}\right]\left[\frac{1}{1-\beta^{2i}}\right]$$

- Clearly, the variance of y_t is positively affected by α_0, α_1 and β.
- We can extend the ARCH model by including more lags

$$e_t = v_t\sqrt{\alpha_0 + \sum_{i=1}^{q}\alpha_i e_{t-i}^2}. \qquad (20)$$

3. GARCH MODELS

- Bollerslev (1986) extended the original work by Engle allowing the conditional variance to be an ARMA process

$$e_t = v_t \sqrt{h_t},$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i e_{t-i}^2 + \sum_{i=1}^{p} B_i h_{t-i}$$

- Given that v_t is a white noise process

$$E_{t-1}(e_t) = E(v_t \sqrt{h_t}) = 0.$$

3. GARCH MODELS

- The expected value of the error term is

$$E_{t-1}(e_t) = E(v_t \sqrt{h_t}) = 0.$$

- We denote this model as GARCH(1,1) and generalizes an ARCH(p) model. GARCH models are often more convenient as they have a much more simple representation.

3. GARCH MODELS

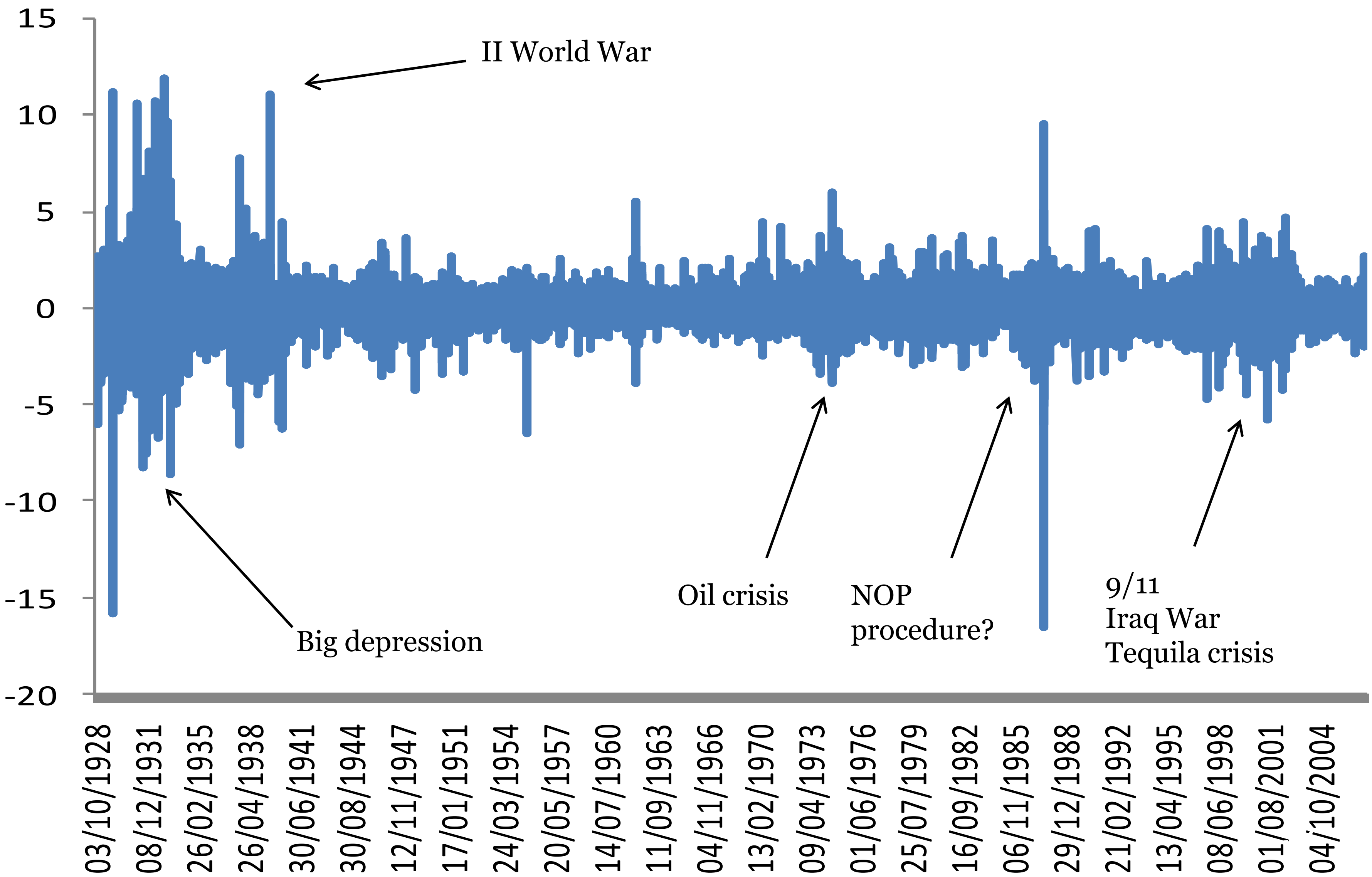
- In practice

1) We estimate an ARIMA model (or any other linear model) such that residuals are white noise.
2) The square values of the residuals should help to identify the GARCH structure

$$E_{t-1}e_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i e_{t-i}^2 + \sum_{i=1}^{p} B_i h_{t-i} \qquad (25)$$

1) Then, we estimate the whole model.

Dayly returns of the Dow-Jones index



4. AN EXAMPLE

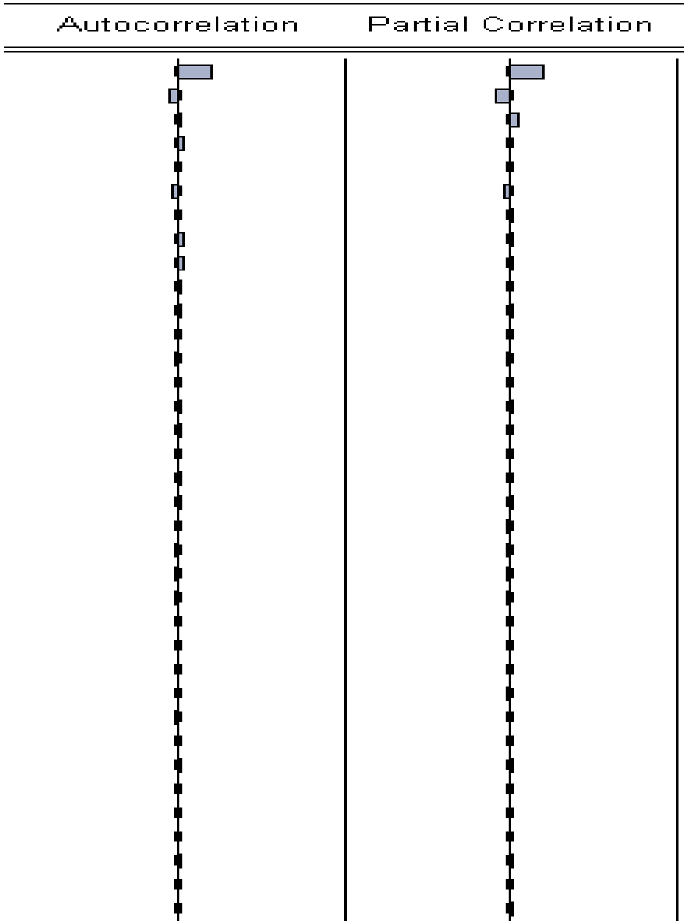
- We have almost 20,000 observations of the dayly returns of the Dow-Jones index 1928-2007.
- Series seem very erratic and it is difficult to propose a stochastic structure for the mean. The correlogram of the series indicates that a MA(1) model with drift could be proposed.
- This constant can be interpreted as an estimation of the unconditional dayly return.

EViews - [Series: DJ Workfile: UNTITLED\Untitled]

File Edit Object View Proc Quick Options Window Help

View Proc Object Properties Print Name Freeze Sample Genr Sheet Stats Ident Line Bar

Date: 11/04/07 Time: 12:07
Sample: 1 19859
Included observations: 19859



	AC	PAC	Q-Stat	Prob
1	0.203	0.203	819.91	0.000
2	-0.049	-0.094	866.67	0.000
3	0.019	0.051	873.83	0.000
4	0.027	0.008	888.70	0.000
5	0.003	0.000	888.95	0.000
6	-0.033	-0.033	910.59	0.000
7	-0.002	0.012	910.67	0.000
8	0.027	0.021	925.60	0.000
9	0.026	0.019	938.77	0.000
10	0.012	0.007	941.43	0.000
11	0.013	0.011	944.68	0.000
12	-0.001	-0.008	944.69	0.000
13	-0.011	-0.009	947.09	0.000
14	-0.007	-0.003	948.06	0.000
15	0.015	0.017	952.30	0.000
16	0.014	0.007	956.35	0.000
17	0.005	0.003	956.76	0.000
18	0.008	0.007	958.16	0.000
19	0.014	0.009	962.29	0.000
20	-0.004	-0.010	962.57	0.000
21	-0.015	-0.010	967.31	0.000
22	-0.024	-0.020	978.73	0.000
23	-0.020	-0.013	986.54	0.000
24	0.001	0.006	986.56	0.000
25	0.000	-0.002	986.56	0.000
26	0.006	0.008	987.32	0.000
27	-0.005	-0.010	987.85	0.000
28	-0.010	-0.007	989.70	0.000
29	0.002	0.004	989.76	0.000
30	0.018	0.019	996.50	0.000
31	0.003	-0.002	996.70	0.000
32	-0.002	0.003	996.78	0.000
33	-0.004	-0.006	997.09	0.000
34	0.015	0.016	1001.9	0.000
35	0.005	-0.004	1002.3	0.000
36	0.005	0.009	1002.7	0.000

Dependent Variable: DJ

Method: Least Squares

Date: 11/04/07 Time: 12:01

Sample: 1 19859

Included observations: 19859

Convergence achieved after 7 iterations

Backcast: 0

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.025145	0.008453	2.974579	0.0029
MA(1)	0.240245	0.006889	34.87488	0

R-squared	0.048717	Mean dependent var	0.025163
Adjusted R-squared	0.048669	S.D. dependent var	0.984786
S.E. of regression	0.960523	Akaike info criterion	2.757424
Sum squared resid	18320.17	Schwarz criterion	2.758219
Log likelihood	-27377.84	F-statistic	1016.909
Durbin-Watson stat	2.026979	Prob(F-statistic)	0

EViews - [Series: RESID Workfile: UNTITLED\Untitled]

File Edit Object View Proc Quick Options Window Help

View Proc Object Properties Print Name Freeze Sample Genr Sheet Stats Ident Line Bar

Date: 11/04/07 Time: 12:01
Sample: 1 19859
Included observations: 19859

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.014	-0.014	3.6723	0.055
		2	-0.051	-0.051	55.405	0.000
		3	0.026	0.025	68.802	0.000
		4	0.020	0.018	76.550	0.000
		5	0.007	0.010	77.475	0.000
		6	-0.035	-0.033	101.16	0.000
		7	0.001	-0.000	101.17	0.000
		8	0.023	0.019	111.38	0.000
		9	0.019	0.022	118.88	0.000
		10	0.004	0.008	119.26	0.000
		11	0.012	0.014	122.11	0.000
		12	-0.001	-0.003	122.14	0.000
		13	-0.009	-0.009	123.68	0.000
		14	-0.008	-0.008	124.98	0.000
		15	0.014	0.014	128.83	0.000
		16	0.011	0.011	131.15	0.000
		17	0.001	0.003	131.17	0.000
		18	0.005	0.005	131.64	0.000
		19	0.014	0.012	135.69	0.000
		20	-0.005	-0.006	136.12	0.000
		21	-0.010	-0.008	138.23	0.000
		22	-0.018	-0.018	144.41	0.000
		23	-0.017	-0.019	150.11	0.000
		24	0.005	0.003	150.70	0.000
		25	-0.003	-0.003	150.84	0.000
		26	0.008	0.009	152.07	0.000
		27	-0.005	-0.006	152.57	0.000
		28	-0.008	-0.009	153.93	0.000
		29	-0.001	-0.002	153.94	0.000
		30	0.019	0.019	160.85	0.000
		31	-0.001	0.001	160.88	0.000
		32	-0.000	0.004	160.88	0.000
		33	-0.008	-0.009	162.10	0.000
		34	0.017	0.015	168.04	0.000
		35	-0.001	-0.002	168.04	0.000
		36	0.005	0.009	168.62	0.000

4. An Example

- Residuals of the model seem to be white noise.
- Can we do something else to understand the dynamic of the Dow Jones index?
- We look at the square values of the residuals.
- If the residuals are white noise with homoskedastic variance, their square values would not have any correlation.

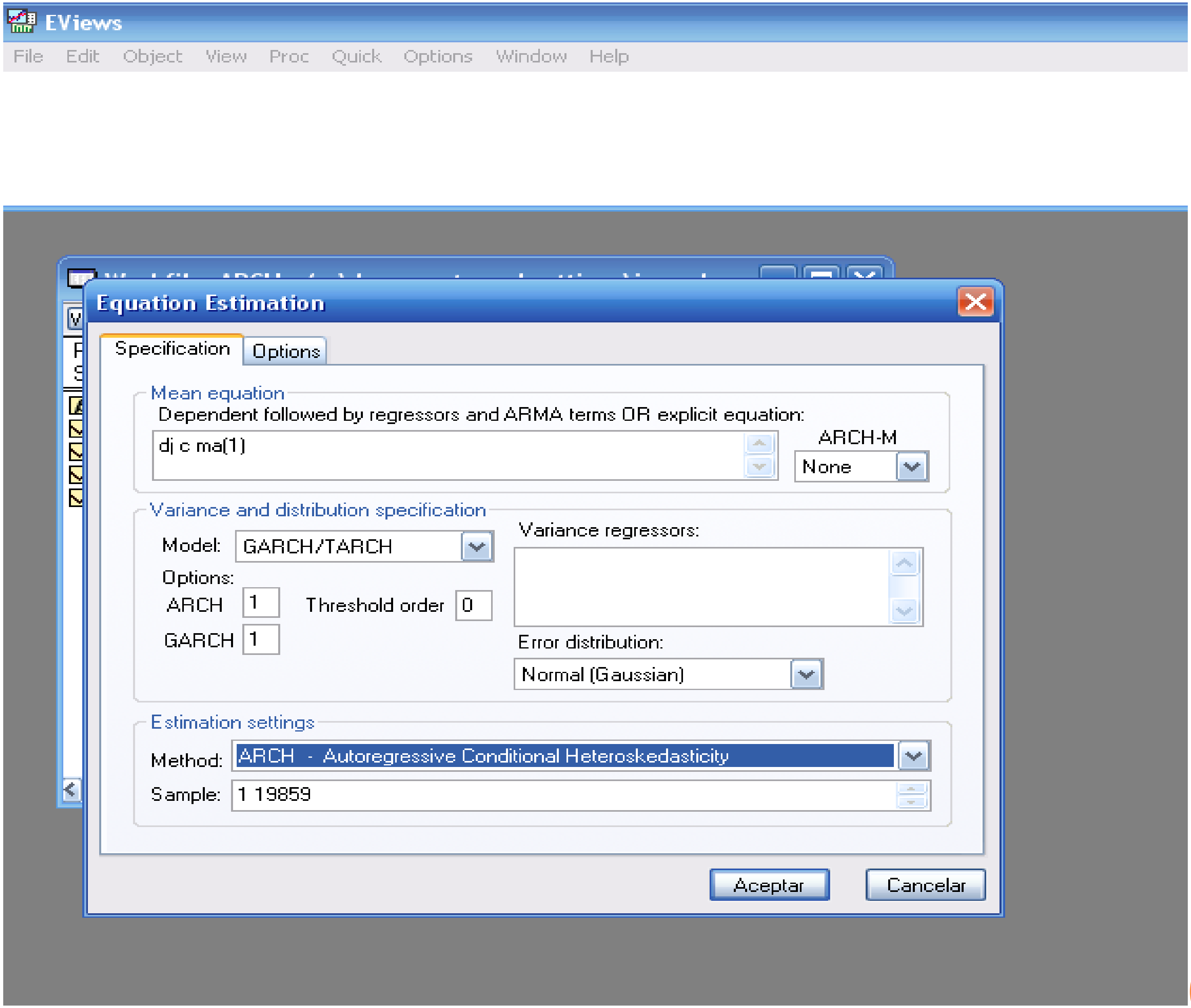
EViews - [Series: E2 Workfile: UNTITLED\Untitled]

File Edit Object View Proc Quick Options Window Help

View Proc Object Properties Print Name Freeze Sample Genr Sheet Stats Ident Line Bar

Date: 11/04/07 Time: 12:21
Sample: 1 19859
Included observations: 19859

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.301	0.301	1796.0	0.000
		2	0.302	0.232	3606.0	0.000
		3	0.253	0.132	4879.9	0.000
		4	0.196	0.053	5642.1	0.000
		5	0.217	0.092	6573.8	0.000
		6	0.185	0.053	7252.0	0.000
		7	0.169	0.036	7816.2	0.000
		8	0.151	0.022	8269.1	0.000
		9	0.142	0.026	8671.2	0.000
		10	0.147	0.039	9101.9	0.000
		11	0.124	0.011	9405.6	0.000
		12	0.115	0.009	9670.6	0.000
		13	0.116	0.020	9936.0	0.000
		14	0.103	0.010	10146.	0.000
		15	0.121	0.035	10437.	0.000
		16	0.093	-0.000	10610.	0.000
		17	0.098	0.013	10801.	0.000
		18	0.093	0.012	10974.	0.000
		19	0.091	0.013	11137.	0.000
		20	0.101	0.024	11340.	0.000
		21	0.091	0.012	11504.	0.000
		22	0.071	-0.012	11603.	0.000
		23	0.083	0.014	11740.	0.000
		24	0.104	0.043	11956.	0.000
		25	0.118	0.046	12231.	0.000
		26	0.088	-0.004	12387.	0.000
		27	0.113	0.032	12639.	0.000
		28	0.088	0.001	12793.	0.000
		29	0.106	0.026	13018.	0.000
		30	0.079	-0.015	13143.	0.000
		31	0.099	0.023	13337.	0.000
		32	0.110	0.034	13579.	0.000
		33	0.101	0.016	13780.	0.000
		34	0.093	-0.001	13953.	0.000
		35	0.071	-0.019	14053.	0.000
		36	0.065	-0.011	14137.	0.000


EViews
File Edit Object View Proc Quick Options Window Help
Equation Estimation
Specification
Options
Mean equation
Dependent followed by regressors and ARMA terms OR explicit equation:
dj c ma(1)
ARCH-M
None
Variance and distribution specification
Model: GARCH/TARCH
Options:
ARCH 1
Threshold order 0
GARCH 1
Variance regressors:
Error distribution:
Normal (Gaussian)
Estimation settings
Method: ARCH - Autoregressive Conditional Heteroskedasticity
Sample: 1 19859
Aceptar
Cancelar

Dependent Variable: DJ
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 11/06/07 Time: 11:42
Sample: 1 19859
Included observations: 19859
Convergence achieved after 20 iterations
MA backcast: 0, Variance backcast: ON
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.048443	0.005731	8.452082	0
MA(1)	0.284062	0.00699	40.63947	0
	Variance Equation			
C	0.008032	0.000521	15.40847	0
RESID(-1)^2	0.095367	0.002157	44.22072	0
GARCH(-1)	0.897253	0.002552	351.6308	0

R-squared	0.046619	Mean dependent var	0.025163
Adjusted R-squared	0.046427	S.D. dependent var	0.984786
S.E. of regression	0.961654	Akaike info criterion	2.248894
Sum squared resid	18360.56	Schwarz criterion	2.250882
Log likelihood	-22325.39	F-statistic	242.7108
Durbin-Watson stat	2.109756	Prob(F-statistic)	0

Inverted MA Roots -0.28

5. Models for Risk. ARCH-M

- Holt and Aradyula (1990) use data from the chicken market to estimate the following equation (Cobweb model)

$$q_t = a_0 + a_1 p_t^e - a_2 h_t - a_3 pfeed_{t-1} + a_4 hatch_{t-1} + a_5 q_{t-1} + e_t$$

where

p_t^e is the expected price of chicken one period ahead.

h_t is the expected variance of chicken prices. Este es el punto de interés.

$pfeed_t$ is the price of chicken food.

$Hatch_t$ is the price of hatch

5. Models for Risk. ARCH-M

- Estimation procedure
  - We estimate an ARIMA model for prices and check that variance of residuals is heteroskedastic.
  - We estimate a GARCH model for the residuals.
  - In incorporate the estimation of the conditional variance of the residuals into the supply equation.

5. MODELS FOR RISK. ARCH-M

- Engle, Lilien y Robins (1987) extend the ARCH methodology to allow the mean to depend on the variance of the process.

- This is especially interested to measure the impact of risk on the returns of different investments.

Dependent Variable: DJ
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 11/06/07 Time: 11:51
Sample: 1 19859
Included observations: 19859
Convergence achieved after 21 iterations
MA backcast: OFF, Variance backcast: OFF
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

	Coefficient	Std. Error	z-Statistic	Prob.
@SQRT(GARCH)	0.061029	0.022443	2.71923	0.0065
C	0.014631	0.0136	1.075837	0.282
MA(1)	0.283761	0.007018	40.43368	0
	Variance Equation			
C	0.008106	0.000528	15.33851	0
RESID(-1)^2	0.095609	0.002194	43.57805	0
GARCH(-1)	0.896862	0.002577	347.9985	0

R-squared	0.0453	Mean dependent var	0.025163
Adjusted R-squared	0.04506	S.D. dependent var	0.984786
S.E. of regression	0.962343	Akaike info criterion	2.248627
Sum squared resid	18385.96	Schwarz criterion	2.251013
Log likelihood	-22321.74	F-statistic	188.4043
Durbin-Watson stat	2.105777	Prob(F-statistic)	0

Inverted MA Roots	-0.28

6. PROPERTIES OF GARCH(1,1) MODELS

- When we estimate an ARCH or GARCH model, two related equations are considered

$$y_t = a_0 + \beta x_t + e_t$$

$$e_t = v_t\left(\alpha_0 + \sum_{i=1}^{q}\alpha_i e_{t-i}^2 + \sum_{i=1}^{p} B_i h_{t-i}\right)^{0.5}$$

- Where x_t is an exogenous variable.
- Given that

$$e_t^2 = v_t^2 h_t$$

and

$$Ev_t^2 = 1 \qquad E_{t-1}e_t^2 = h_t$$

6. PROPERTIES OF GARCH(1,1) MODELS

- The unconditional mean of e_t is

$$E\ e_t = E\left(v_t (h_t)^{1/2}\right) = 0$$

- The unconditional variance of e_t is

$$E\ e_t^2 = \frac{\alpha_0}{1 - \alpha_1 - B}$$

- More generally, the unconditional variance of e_t in a GARCH(p,q) process is finite if

$$1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} B_i > 0$$

6. Properties of GARCH(1,1) models

- The FAC of e_t is

$$Ee_t e_{t-j} = E\left[v_t (h_t)^{1/2} v_{t-j} (h_{t-j})^{1/2}\right] = 0 \qquad (33)$$

For j>0.

- The error term in GARCH model is sequentially uncorrelated. However, the square of the errors are correlated.
- Indeed, the square of the error term follows an ARMA model.
- Parameters must be restricted in order to ensure that the square of residuals never takes a negative value.

6. Properties of GARCH(1,1) models

- We can not look at the AIC and SBC in order to select an ARCH or GARCH model.
- To see this, we look at the definition of the AIC

$$AIC = T\ln\begin{pmatrix} \text{sum} & \text{of} & \text{square} \\ \text{residuals} & & \end{pmatrix} + 2n \qquad (34)$$

- In order to check whether an ARCH(2) model is a better specification than a GARCH(1,1) model, we look at

$$RSS' = \sum_{t=1}^{T} (e_t^2 - h_t)^2 \qquad (35)$$

6. Properties of GARCH(1,1) models

- In order to propose an analogous measure of AIC and SCB for ARCH and GARCH, we consider the maximum value of the likelihood function:

$$L = -\sum_{t=1}^{T}\left[\ln(h_t) + e_t^2/h_t\right]$$

- We can include a penalization for additional parameters and propose the following measures

$$AIC' = -\ln L + 2n$$
$$SBC' = -\ln L + n\ln(T)$$

7. Diagnosis and forecast

- Diagnosis
  - Standarized residuals should have zero mean and unit variance.
  - Standarized residuals should be serially uncorrelated. Check the correlograms.
- Forecast

$$h_{t+1} = \alpha_0 + \alpha_1 e_t^2 + B_1 h_t$$

Convergence condition:

$$E_t(h_{t+j}) = \alpha_0 + \left[1 + (\alpha_1 + B_1) + (\alpha_1 + B_1)^2 + \ldots + (\alpha_1 + B_1)^{j-1}\right] + (\alpha_1 + B_1)^j h_t \qquad (38)$$

$$(\alpha_1 + B_1) < 1$$

Smooth transition regression modeling

72

The standard smooth transition regression (STR) model is defined as follows

$$y_t = \phi' z_t + \theta' z_t G(\gamma, c, s_t) + u_t$$
$$= \{\phi + \theta G(\gamma, c, s_t)\}' z_t + u_t$$

where $z_t = (w_t', x_t')'$ is a vector of explanatory variables, $w_t' = (1, y_{t-1}, \dots, y_{t-p})'$ and $x_t = (x_{1t}, \dots, x_{kt})'$ is a vector of exogenous variables.

$$u_t \sim iid(0, \sigma^2)$$

Transition function $G(\gamma, c, s_t)$ is a bounded function of the continuous variable s_t, continuous everywhere in the parameter space for any value s_t, γ is the slope parameter, and $c = (c_1, \ldots, c_k)'$ which is a vector of location parameters $c_1 \leq \cdots \leq c_k$.

The transition function can be assumed to be a general logistic function

$$G(\gamma, c, s_t) = \left(1 + exp\left\{-\gamma \prod_{k=1}^{K} (s_t - c_k)\right\}\right)^{-1}, \gamma > 0$$

where γ is an identifying restriction.

The most common choices for K are $K = 1$ and $K = 2$.

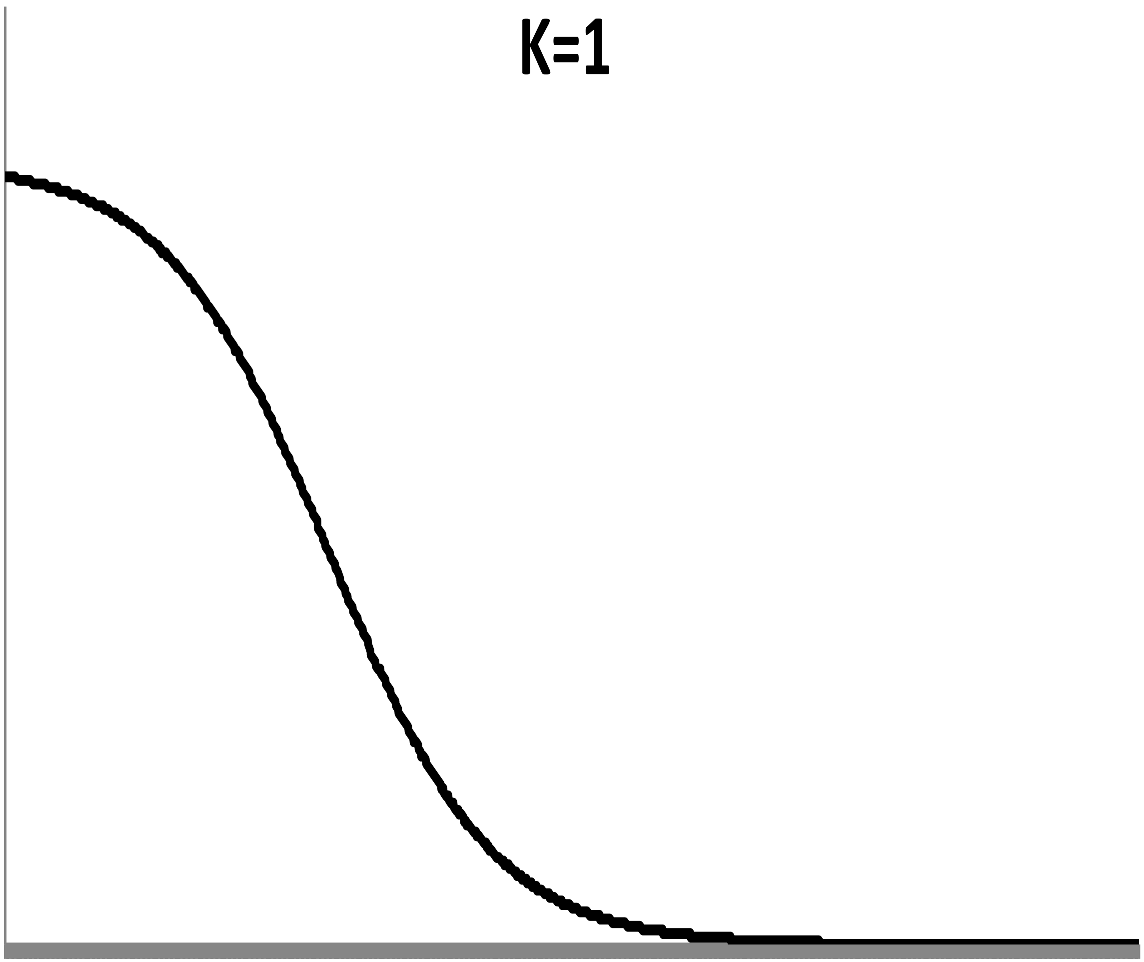
For $K = 1$, the parameters $\phi + \theta G(\gamma, c, s_t)$ change monotonically as a function of s_t, from ϕ to $\phi + \theta$.

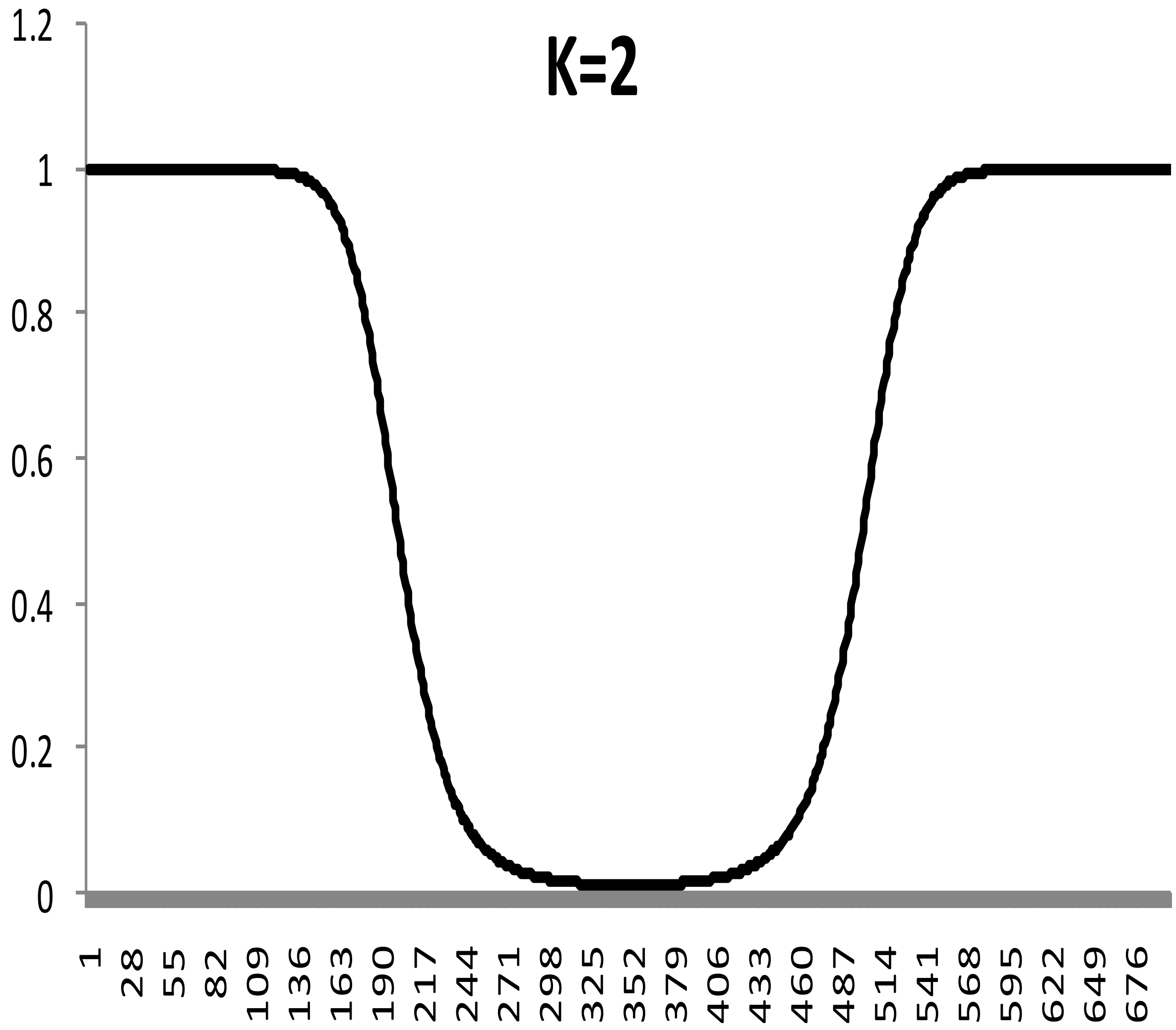
For $K = 2$, they change symmetrically around the midpoint $\frac{(c_1+c_2)}{2}$ where this logistic function attains its minimum value. This minimum value lies between zero and ½. It reaches one when $\gamma \to \infty$ and equals ½ when $s_t = c_1$ or $s_t = c_2$ and $\gamma < \infty$.

The LSTR model with $K = 1$ is capable to characterize asymmetric behavior. As an example, suppose that s_t measures the phase of the business cycle. Then the LSTR1 model can describe processes whose dynamic properties are different in expansions from what they are in recessions, and the transition from one extreme regime the other is smooth.

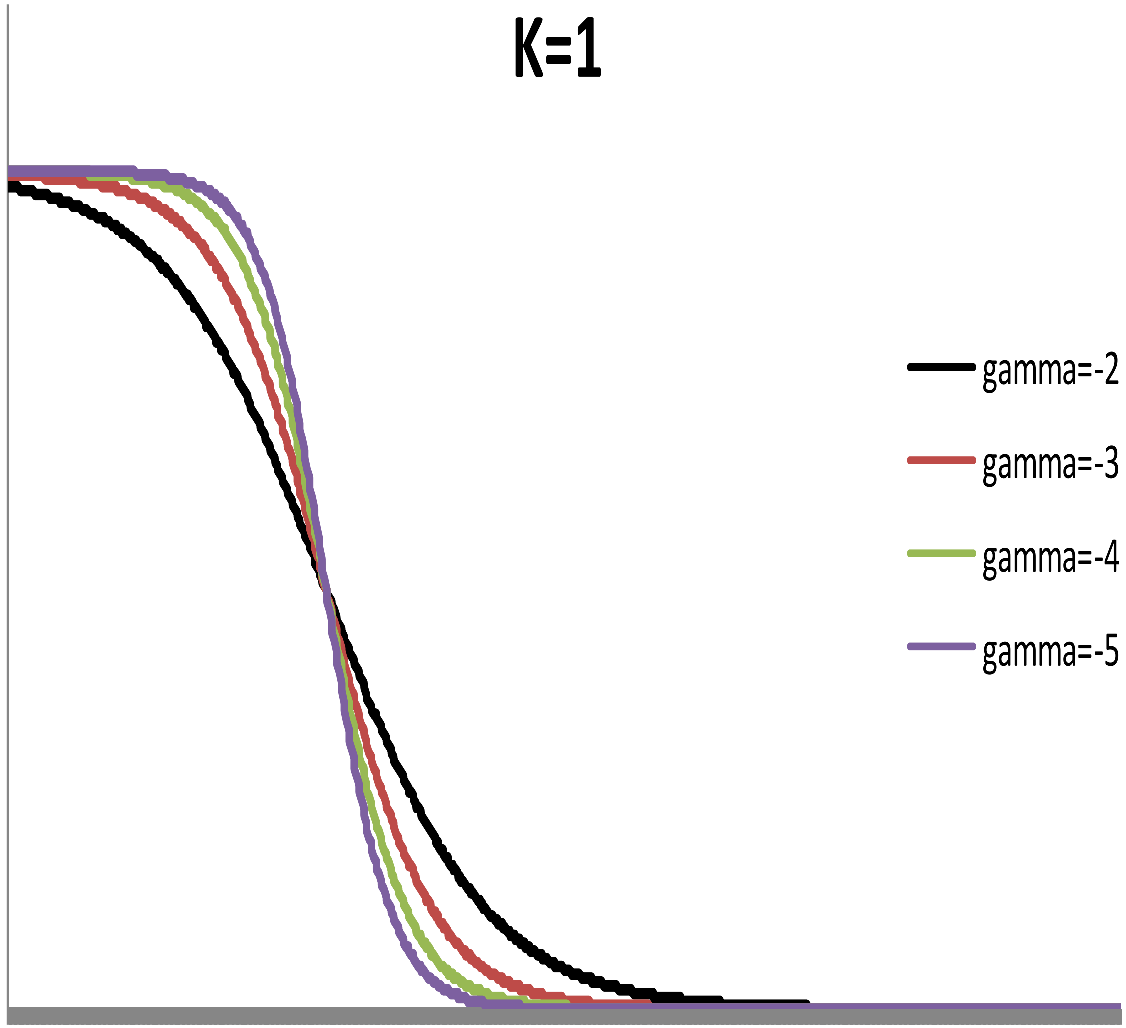
On the other hand, the LSTR2 model is appropiate in situations in which the local dynamic behavior of the process is similar at both large and small values of s_t and different in the middle. For example, a nonlinear equilibrium correction in which the strength of attraction varies nonlinearly as a funtion of the size of the deviation from the equilibrium. See Öcal and Osborn (2000) and van Dijk and Franses (1999) for works on this issue.

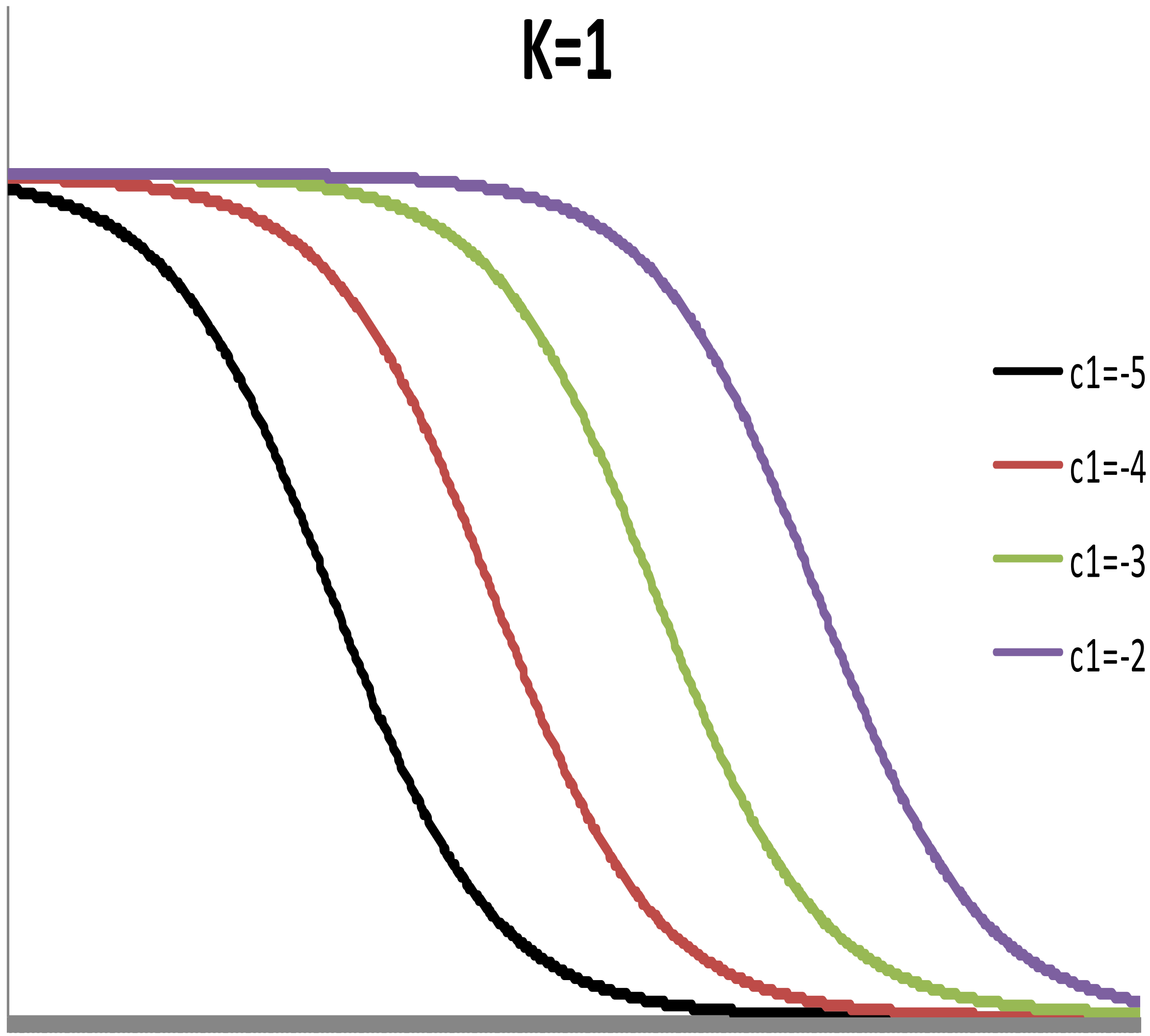
Slope parameter γ controls the slope and c_1 and c_2 the localization of the transition function.


K=1



K=2
1.2
1
0.8
0.6
0.4
0.2
0
1
28
55
82
109
136
163
190
217
244
271
298
325
352
379
406
433
460
487
514
541
568
595
622
649
676


K=1
gamma=-2
gamma=-3
gamma=-4
gamma=-5


K=1
c1=-5
c1=-4
c1=-3
c1=-2

- When $\gamma = 0$, the transition function $G(\gamma, c, s_t) \equiv 1/2$ and thus the STR model nests the linear model.

- When $\gamma \to \infty$, the LSTR1 model approaches the switching regression model with two regimes that have equal variances.

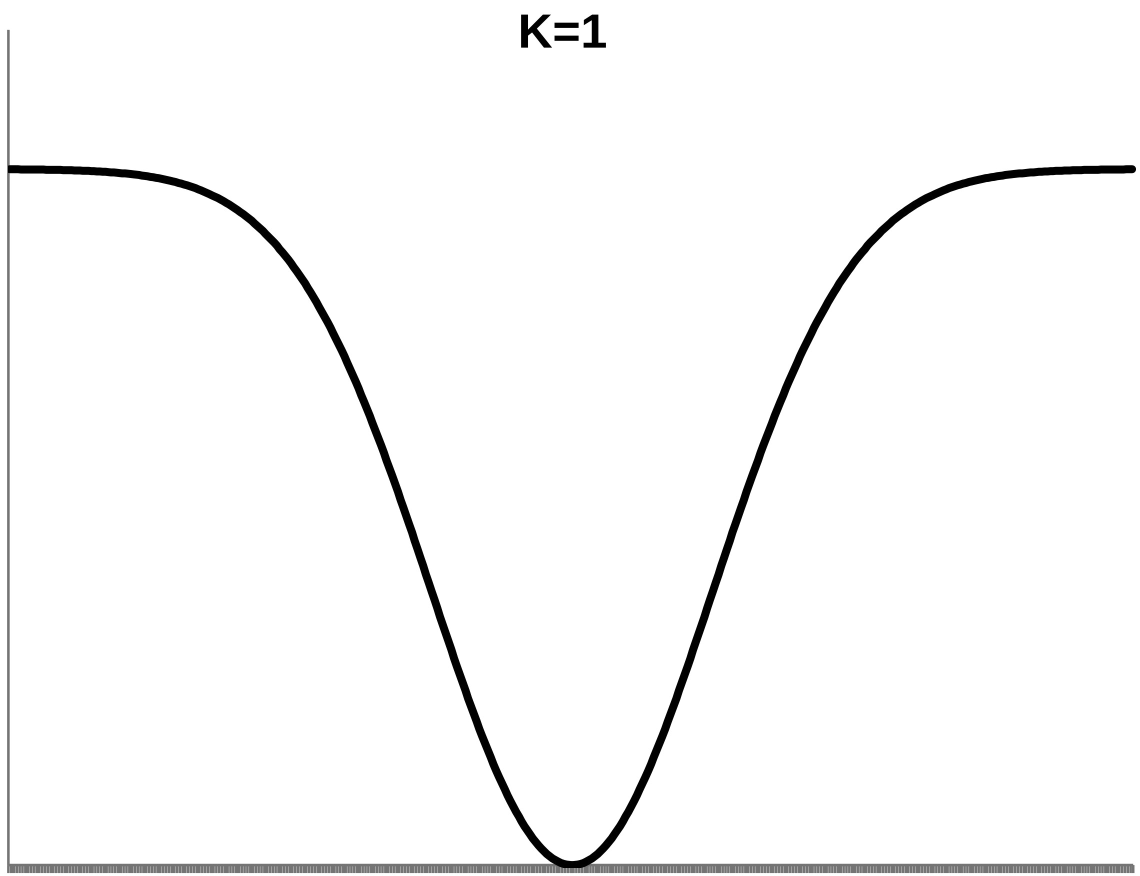
- When $\gamma \to \infty$ in the LSTR2 model, the result is another switching regression model with three regimes such that the outer regimes are identical to the midregime and the midregime is different to the other two.

An alternative to the LSTR2 model is the so-called exponential STR model with the transition function

$$G(\gamma, c, s_t) = 1 - exp\{-\gamma(s_t - c_1^*)^2\}, \qquad \gamma > 0.$$

This function is symmetric around $s_t = c_1^*$ and has, at low and moderate values of the slope parameter γ, approximately the same shape, albeit a different minimum value (zero). Because this function contains one parameter less than the LSTR2 model, it can be regarded as a useful alternative to the corresponding logistic transition function.

It has a drawback, however. When $\gamma \to \infty$ the function becomes particularly linear, for the transition function equals zero at $s_t = c_1^*$ and unity elsewhere. The ESTR model is not a good approximation to the LSTR2 model when γ in the latter is large and $c_2 - c_1$ is at the same time not close to zero.


K=1

In practice, the transition variable s_t is a stochastic variable and very often an element of z_t.

It can also be a combination of several variables. In some cases, it can be a difference of an element of z_t; see Skalin and Teräsvista (2002) for an univariate example.

A special case, $s_t = t$, yields a linear model with deterministically changing parameters.

When x_t is absent from the specification and $s_t = y_{t-d}$ or $s_t = \Delta y_{t-d}$, $d > 0$, the STR model becomes a univariate smooth transition autoregressive model; see Teräsvirta (1994) for some discussion.

Model specification

The specification stage entails two phases.

1) the linear model forming the starting point is subject to linearity test and;
2) the type of STR model (LSTR1 or LSTR2) is selected.

Economic theory may give an idea of which variables should include in the linear model but may not be particularly helpful in specifying the dynamic structure of the model.

However, If the model is a purely autoregressive one, it may be advisible not to create gaps in the lag structure by omitting lag shorter than the maximum lag selected for the model. The reason is that such omissions may reduce the power of the linearity tests. Models for strongly seasonal time series constitute an obvious exception to this rule.

Linearity is tested against an STR model with a predetermined transition variable. If economic theory is not explicit about this variable , the test is repeated for each variable in the predetermined set of potential transition variables, which is usually a subset of the elements in z_t. The purpose of these tests is twofold:

1) They are used to test linearity against different directions in the parameter space. If no rejection of the null hypothesis occur, the model builder accepts the linear model and does not proceed with STR models.

2) The test results are used for model selection. If the null hypothesis is rejected for at least one of the models, the model against which the rejection, measured in the p-value, is strongest is chosen to be the STR model to be estimated.

- Testing linearity; Luukonen, Saikkonen and Teräsvirta (1988) and Teräsvirta (1994, 1998). The STR model shares with many other nonlinear models the propertiy that the model is only identified under the alternative.

- The ensuing identification problem in testing linearity can, in the STR context, be circumvented by approximating the logistic transition function by a Taylor expansion around the null hypothesis $\gamma = 0$. It is customary to assume $K = 1$ and use the third-order Taylor approximation; see Teräsvirta (1998). The resulting test has power against both the LSTR1 ($K = 1$) and LSTR2 ($K = 2$) models.

Assume now that the transmission variable s_t is an element in z_t and let $z_t = (1, \tilde{z}_t)'$, where $\tilde{z}_t$ is an $(mx1)$ vector. The approximation yields, after merging terms and reparameterizing, the following auxiliary regression:

$$y_t = \beta_0{}' z_t + \sum_{j=1}^{3} \beta_j{}' \tilde{z}_t s_t^j + u_t^*$$

where $u_t^* = u_t + R_3(\gamma, c, s_t)\theta' z_t$ with the remainder $R_3(\gamma, c, s_t)$. The null hypothesis is $H_0\colon \beta_1 = \beta_2 = \beta_3$ because each β_j, $j = 1,2,3$, is of the form $\gamma\tilde{\beta}_j$, where, $\tilde{\beta}_j \neq 0$ is a function of θ and c. This is a linear hypothesis in a linear (in parameters) model. Because $u_t^* = u_t$ under the null hypothesis, the asymptotic distribution theory is not affected if an LM-type test is used.

- The asymptotic distribution theory of the resulting χ^2-test requires the existence of $Es_t^6\tilde{z}_t\tilde{z}_t{}'$. If the model is a univariate STAR model, this is equivalent to requiring $Ey_t^8 < \infty$. This assumption naturally implies restrictions on β_0.

- The test statistic has an asymptotic χ^2-distribution with 3m degrees of freedom when the null hypothesis is valid. But then, the χ^2-statistic can be severely size-distorted in small and even moderate samples. The corresponding F-statistic is recommended instead. It has an approximate F-distribution with 3m and T-4m-1 degrees of freedom under the null hypothesis. JMulti always use the F-version of the test.

In building STR models, the test is applied as follows.

1) Select a set of potential transition variables $S = \{s_{1t}, \dots, s_{kt}\}$. It may contain the same elements as $\tilde{z}_t$, but economic theory and other considerations may restiric the set or suggest adding other variables.

2) After defining S, perform the test using each element in S in turn as the transition variable.

3) If the null hypothesis is rejected for several restriction variables, select the one for which the p-value of the test is minimized. The logic behind this suggestion is that the rejection of the null hypothesis is stronger against the correct alternative than other alternatives. However, if several small p-values are close to each other, it may be useful to proceed by estimating the corresponding STR models and leaving the choice between them to the evaluation stage; see Teräsvirta (1998).

Choosing the type of the model

- When the linearity has been rejected and a transition variable subsequently selected, the next step will be to choose the model type. The available choices are $K = 1$ and $K = 2$ in the logistic function type.

- Alternatively, instead of $K = 2$, one may in some situations use the exponential transition function. This can be the case when, for example, the sequences of estimates of c_1 and c_2 in the iterative estimation converge toward the same value.

- The choice between different models can be based on the auxiliary regression for the linear test. In the special case $c = 0$, it can be shown that $\beta_2 = 0$ when the model is an LSTR1 model, whereas $\beta_1 = \beta_3 = 0$ when the model is an LSTR2 or ESTR model.

This suggests the following short test sequence:

1) Test the null hypothesis $H_{04}: \beta_3 = 0$.
2) Test $H_{03}: \beta_2 = 0 \backslash \beta_3 = 0$.
3) Test $H_{02}: \beta_1 = 0 \backslash \beta_2 = \beta_3 = 0$.

If the test of H_{03} yields the strongest rejection measured in the p-value, choose the LSTR2 or ESTR model. Otherwise, select the LSTR1 model. All three hypothesis can simultaneously be rejected at a conventional significance level such as 0.05 or 0.01; that is why the strongest rejection counts.

It is also possible to fit both an LSTR1 and LSTR2 model to the data and make the choice between them at the evaluation stage. This is a sensible way of proceeding if the sequence does not provide a clear-cut choice between the two alternatives.

Reducing the size of the model

As in linear models, the model builder often wants to reduce the size of the model by eliminating redundant variables.

Eliminating an element in z_t such as z_{jt} requires the restriction $\phi_j = \theta_j = 0$. Unlike the situation for linear models, two other types of exclusion restrictions are of interest.

1) $\phi_j = 0$. This restriction limits the combined coefficient of z_{jt} to zero for $G(\gamma, c, s_t) = 0$ so that z_{jt} does not contribute in that regime.
2) A mirror image of this restriction is is $\phi_j = -\theta_j$, which limits the combined coefficient to zero when $G(\gamma, c, s_t) = 1$.

Thus, in reducing the number of parameters, restrictions $\phi_j = 0$ and $\phi_j = -\theta_j$ should both be considered. Naturally, restricting z_{jt} to only appear linearly ($\theta_j = 0$) has to be considered as well.

ESTIMATION

- Initial values: the parameters of the STR model are estimated using conditional maximum likelihood. The log-likelihood is maximized numerically and JMulTi uses the iterative BFGS algorithm, Hendry (1995).
- Finding good starting-values for the algorithm is important. One way of obtaining them is the following. When γ and c in the transition function are fixed, the STR model is linear. This suggests constructing a grid and select the parameters that minimize the sum of squared residuals.

When constructing the grid, note that γ is not a scale-free parameter. The exponent of the transition function is therefore standarized by dividing it by the Kth power of the sample standard deviation of the transition variable s_t, which we will call $\hat{\sigma}_s$. The transition function becomes

$$G(\gamma, c, s_t) = \left(1 + exp\left\{-{}^{\gamma}/_{\hat{\sigma}_s^K} \prod_{k=1}^{K} (s_t - c_k)\right\}\right)^{-1}, \gamma > 0$$

A specific numerical problem exists in the estimation of STR models. It is present when γ is large and the model is consequently close to a switching regression model. This makes the estimation of γ difficult in small samples because determining the curvature of the transition function requires many observations in the neighborhood of c $(K = 1)$ and c_1 and $c_2 (K = 2)$. This lack of information manifests itself in the standard deviation estimate of $\hat{\gamma}$, which becomes large.

EVALUATION. TESTING THE STR MODEL

Eitrheim and Teräsvirta (1996) and Teräsvirta (1998) have considered misspecification testing in ST(A)R models

Test of no error autocorrelation: is a special case of a general test discussed in Godfrey(1998). Assume that $M(z_t;\psi)$ is at least twice continuously differentiable with respect to the parameters wverywhere in the sample space and that

$$y_t = M(z_t;\psi) + u_t, t = 1, \dots, T$$

where $u_t = \alpha' v_t + \varepsilon_t$ with $\alpha = (\alpha_1, \dots, \alpha_q)'$, $v_t = (u_{t-1}, \dots, u_{t-q})'$ and $\varepsilon_t \sim iidN(0, \sigma^2)$. The null hypothesis of no error autocorrelation against the alternative of autocorrelation of at most order q in u_t is $\alpha = 0$.

Briefly, the test consists of regressing the residuals $\tilde{u}_t$ of the estimated STR model on the lagged residuals $\tilde{u}_{t-1}, \dots, \tilde{u}_{t-q}$ and the partial derivatives of the log-likelihood function with respect to the parameters of the model evaluated at the maximizing value $\psi = \hat{\psi}$. Let n be the number of parameters in the model. Then, the test statistic

$$F_{LM} = \{(SSR_0 - SSR_1)/q\}/\{SSR_1/(T - n - q)\}$$

where SSR_0 is the sum of squared residuals of the STR model and SSR_1 the corresponding sum from the auxiliary regression just described, has an approximate F-distribution with q and $T - n - q$ degrees of freedom under the null hypothesis. The F-version of the test is preferable to the corresponding χ^2-statistic based on the asymptotic distribution theory. When the model is linear, the test collapses in to the test of no serial correlation of Breusch (1978) and Godfrey (1978).

Test of no additive nonlinearity

After an STR model has been fitted to the data, it is important to ask whether the model adequately characterizes the nonlinearity originally found in the data by applying linearity test or to ask whether some nonlinearity remains unmodeled. In the STR framework, a natural alternative to consider in this context is an additive STR model. It can be defined as follows:

$$y_t = \phi' z_t + \theta' z_t G(\gamma_1, c_1, s_{1t}) + \psi' z_t H(\gamma_2, c_2, s_{2t}) + u_t$$

where $H(\gamma_2, c_2, s_{2t})$ is another transition function and $u_t \sim iidN(0, \sigma^2)$. For notational simplicity, assume $H(0, c_2, s_{2t}) = 0$.

Then, the null hypothesis of no additive nonlinarity can be defined as $\gamma_2 = 0$. The model is only identified under the alternative: both ψ and c_2 are nuisance parameters under the null hypothesis. The identification problem can be solved by approximating the transition function H by its Taylor expansion around $\gamma_2 = 0$, merging terms, and reparameterizing. If a third order expansion is assumed, this yields the following auxiliary model:

$$y_t = \beta_0^{'} z_t + \theta^{'} z_t G(\gamma_1, c_1, s_{1t}) + \sum_{j=1}^{3} \beta_j{}' \left(\tilde{z}_t, s_{2t}^{j}\right) + u_t^{*},$$

where $u_t^{*} = u_t + \psi' z_t R_3(\gamma_2, c_2, s_{2t})$.

The null hypothesis is $\beta_1 = \beta_2 = \beta_3 = 0$.

Deriving the LM-type test for testing this hypothesis is straightforward. The only difference compared with the case of testing linearity is that z_t is replaced by the gradient vector $v_t = \left(z_t^{'}, z_t^{'} G(\tilde{\gamma}_1, \tilde{c}_1, s_{1t}), g_t(\tilde{\gamma}), g_t(\tilde{c}_1)\right)^{'}$, where

$$g_t(\tilde{\gamma}) = \partial G(\gamma_1, c_1, s_{1t})/\partial \gamma_1 \,|_{(\gamma_1, c_1) = (\widetilde{\gamma}_1, \widetilde{c_1})}$$

and

$$g_t(\tilde{c}_1) = \partial G(\gamma_1, c_1, s_{1t})/\partial c_1 \,|_{(\gamma_1, c_1) = (\widetilde{\gamma}_1, \widetilde{c_1})}$$

The moment requirement is $Es_t^6 z_t z_t^{'} < \infty$. The test can be restricted to concern only a subvector of ψ by assuming that some elements of ψ equal zero a priori.

This test can in practice be applied in the same way as the linearity test by defining the set S of potential transition variables and carrying out the test against every variable in the set.

As a special case, one may test restrictions imposed on the STR model. Assume that the estimated equation contains the exclusion restriction $\phi^{(0)} = 0$ or, alternatively $\phi^{(0)} = -\theta^{(0)}$, where $\phi^{(0)}$ and $\theta^{(0)}$ are subsets of elements of ϕ and θ.

Consider the first case. The validity of the restriction may be tested after estimating the restricted model by testing the hypothesis $\phi^{(0)} = 0$ in the linear augmented STR equation

$$y_t = \phi^{(1)\prime} z_t^{(1)} + \theta' z_t G(\gamma_1, c_1, s_{1t}) + \phi^{(0)\prime} z_t^{(0)} + u_t$$

Where $z_t = \left(z_t^{(0)'}, z_t^{(1)'}\right)'$.

A corresponding test is available for testing the validity of $\phi^{(0)} = -\theta^{(0)}$.

An alternative parsimonious misspecification test would be the RESET of Ramsey (1969). This test is carried out by testing the hypothesis $\beta_2 = \cdots = \beta_h$ in another linearly augmented equation

$$y_t = \beta_0^{'} z_t + \theta^{'} z_t G(\gamma_1, c_1, s_{1t}) + \sum_{j=1}^{3} \beta_j{}' \tilde{y}_t^j + u_t$$

where $\tilde{y}_t$ is the fitted value of y_t from the estimated STR model. It should be pointed out, however, that, in practice, RESET has turned out to be a very powerful test in detecting misspecification of the STR model.

TESTING PARAMETER CONSTANCY

Parameter nonconstancy may indicate misspecification of the model or genuine change over time in the economic relationship described by the model. Either way, parameter constancy is one of the hypothesis that has to be tested before the estimated model can be used for forecasting or policy simulations.

The alternative to parameter constancy allows smooth transition change in parameters.

We consider the following model

$$y_t = \phi(t)' z_t + \theta(t)' z_t G(\gamma, c, s_t) + u_t \ , \gamma > 0$$

where

$$\phi(t) = \phi + \lambda_\phi H_\phi(\gamma_\phi, c_\phi, t^*)$$

and

$$\theta(t) = \theta + \lambda_\theta H_\theta(\gamma_\theta, c_\theta, t^*)$$

where $t^* = {}^{t}/_{T}$ and $u_t \sim iidN(0, \sigma^2)$.

The null hypothesis of parameter constancy equals $\gamma_\phi = \gamma_\theta = 0$ whereas the alternative H_1: "either $\gamma_\phi > 0$ or $\gamma_\theta > 0$ or both."

The TV-STR model is only identified when $\gamma_\phi, \gamma_\theta > 0$. To circumvent the problem we proceed as before and expand the expressions around the null hypothesis. A first-order Taylor expansion around $\gamma_\psi = 0$, if the order of the logistic function $K = 3$, has the following form after reparameterization:

$$
\begin{aligned}
&T\left(\gamma_\psi, c_\psi, t^*\right) \\
&\qquad = (1/2) + (\gamma_\psi/2)\left\{\delta_0^{(\psi)} + \delta_1^{(\psi)} t^* + \delta_2^{(\psi)} (t^*)^2 + \delta_3^{(\psi)} (t^*)^3\right\} \\
&\qquad + R_1(\gamma_\psi, c_\psi, t^*)
\end{aligned}
$$

for $\psi = \theta, \phi$, where R_1 is the remainder.

If one insert this approximation in the model gets the following nonlinear auxiliary regression:

$$y_t = \beta_0{}'z_t + \sum_{j=1}^{3} \beta_j{}'\{z_t(t^*)^j\} + \sum_{j=1}^{3} \beta_{j+3}{}'\{z_t(t^*)^j\}G(\gamma, c, s_t) + u_t^*$$

Where $\beta_j = 0, j = 1, \dots, 6$, if and only if the null hypothesis $\gamma_\phi = \gamma_\theta = 0$ holds.

It is possible testing subhypotheses by assuming that the parameters not under test are constant; see Teräsvista (1998) for discussion.

The LM-type thest for testing the current null hypothesis is analogous to the ones in the preceding sections. The auxiliary regression now consist of regressing residuals $\tilde{u}_t$ (or y_t) on

$$v_t = [z_t', z_t' t^*, z_t'(t^*)^2, \dots, z_t'(t^*)^3 G(\gamma, c, s_t)]$$

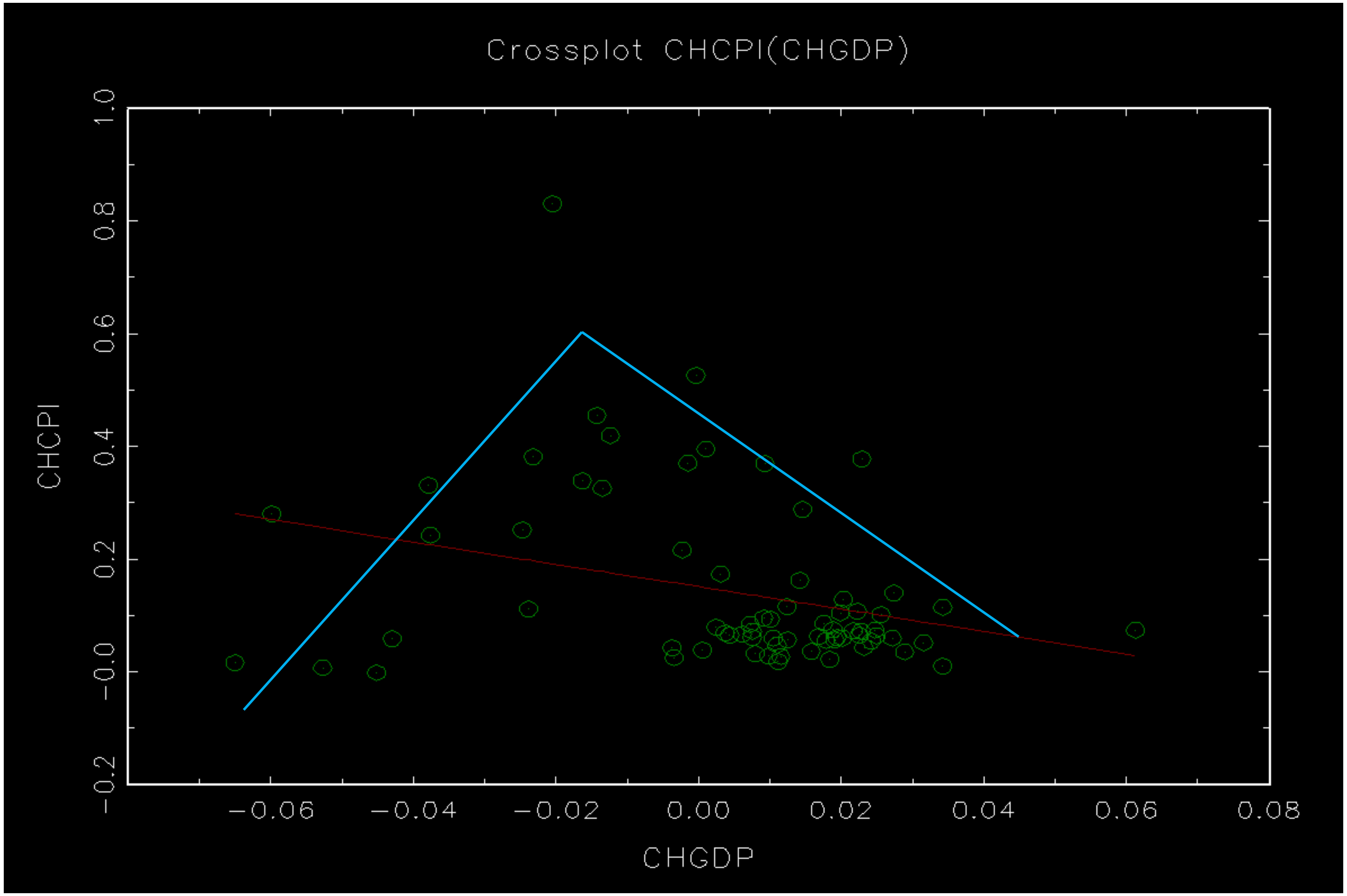
The F-version of the test is recommended because v_t is $(7(m+1)x1)$ vector and the number of degrees of freedom in the χ^2-test would equal $6(m+1)$.

Although the test just discussed may be the most obvious ones to use when an estimated STR model is evaluated, other tests may be useful. One may, for example, test the null hypothesis of no ARCH.

WHAT TO DO WHEN AT LEAST ONE TEST REJECT

- Error autocorrelation indicates misspecification but it is not specific about its nature. The test may not only have power against misspecified dynamics but also against omitted variables.
- Rejecting the null of no additive nonlinearity may suggest adding another STR component to the model. But then, because a rejection as such does not say anything definite about the cause, the idea of extending the model further has to be weighted against other considerations such as the risk of overfitting.

An Example: Relation between the inflation and GDP growth in Chile.



- It seems that when growth is positive, there is a negative relationship between inflation and output but it is positive in period of crisis. This is because we are considering two different historical periods, the seventies were inflation was mainly affected by productivity shocks and the nineties when demand shocks were far more important.

- What k would I choose?

That is, the LSTR1 model can be written as (standard values between brackets)

$$\pi_t = \underset{(0.11)}{0.57} + \underset{(2.51)}{8.71} y_t - \left(\underset{(0.13)}{0.48} + \underset{(3.28)}{10.03} y_t \right) \left(1 + exp \left\{ - \underset{(2.73)}{5.38/\hat{\sigma}_{y_t}} (y_t + \underset{(0.004)}{0.09}) \right\} \right)^{-1} + \hat{u}_t$$

$$Adjusted\ R^2 = 0.40$$

It seems reasonable to consider a model in which parameters change through time (but the specification is worse) p-values of the linear test (using the first model)

Hypothesis	Transition variable y_t
H_0	2.59e-3
H_{04}	7.49e-1
H_{03}	1.98e-3
H_{02}	3.12e-2

Because H02 and H04 are larger than H03 the choice seem K=2.

$$\pi_t = \underset{(0.03)}{0.18} + \underset{(3.48)}{-18.13} y_t$$

$$+ \left(\underset{(0.04)}{0.08} \; \underset{(3.53)}{-17.48} y_t \right) \left(1 + exp \left\{ - \frac{\underset{}{32}}{\underset{(4.30)}{\hat{\sigma}_{y_t}}} \left(y_t + \underset{(0.00)}{0.02} \right) \left(y_t + \underset{(0.00)}{-0.005} \right) \right\} \right)^{-1} + \hat{u}_t$$

$$Adjusted\ R^2 = 0.49$$

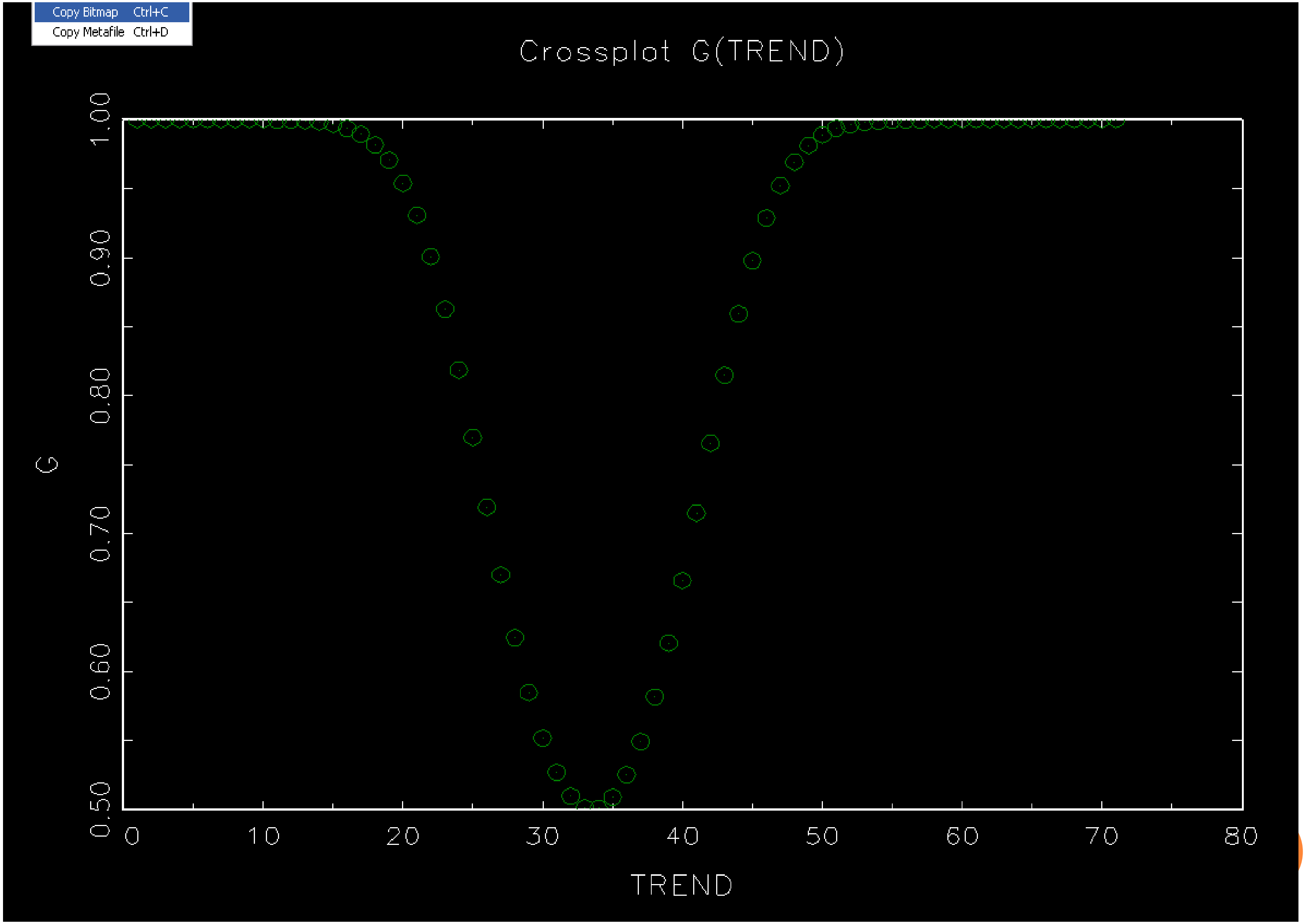
It seems that the constant term is very similar in both stages, so we include this restriction in the estimation

$$\pi_t = \underset{(0.06)}{0.47} + \underset{(2.75)}{-11.5} y_t + \left(\underset{(0.06)}{0.47} \underset{(2.95)}{-10.5} y_t \right) \left(1 + exp \left\{ - \frac{32}{\underset{(4.30)}{\hat{\sigma}_{y_t}}} \left(y_t - \underset{(0.00)}{0.02} \right) \left(y_t + \underset{(0.00)}{0.005} \right) \right\} \right)^{-1}$$

$$+ \hat{u}_t$$

$$Adjusted\ R^2 = 0.76$$

The fit is also better


Copy Bitmap Ctrl+C
Copy Metafile Ctrl+D
Crossplot G(TREND)
G
1.00
0.90
0.80
0.70
0.60
0.50
0
10
20
30
40
50
60
70
80
TREND

FORECASTING

- The STAR model is a nonlinear model, which makes multi-period forecasting from them more complicated than from linear models.
- In forecasting a single period ahead there is no difference between STAR and linear AR models.
- An exposition for obtaining multi-period forecasts can be found in Granger and Teräsvirta (1993).

- An advantage of numerical approximations to the true expectations is that they automatically give a number of point forecasts for each period to be predicted. In fact, what is available is a forecast density, and interval forecasts may be constructed on the basis of them.

- In so doing, one has to remember that the interval forecasts obtained in this way do not account for sampling uncertainty and the intervals are therefore two narrow.

- Sampling uncertainty may be accounted for by numerical techniques.
- Suppose $\hat{\psi}$ is a consistent and asymptotically normal estimator with a positive definite covariance matrix of the parameter vector ψ.
- The forecasts obtained by any of the aforementioned techniques are based on the estimate $\hat{\psi}$.
- We can draw a random sample of size R, say, from the large-sample distribution of the estimator $\hat{\psi}$ and obtain R sets of new estimates.
- Plugging them gives R new models. By repeating the procedure (ii) and (iii) for each of these leads to RM or RB forecasts. The interval forecasts based on these predictions now accommodate the sampling uncertainty.

- This approach is quite computer-intensive.
- In the case of SETAR models it also requires plenty of human resorces.
- We cannot exclude the possibility that some of the R models are unstable, and instability in turn would mess up the forecasts.
- Every model thus has to be checked for stability, which requires human control and an even greater amount of computational resorces.

When we generate a multi-period forecasts with Monte Carlo and bootstrap procedures we have a whole set of forecasts for each period.

A density forecast for each of these periods may be obtained as follows:

1) Arrange the individual forecasts for a given time period in ascending order.
2) Represent this empirical distribution as a histogram and smooth it to obtain an estimate of the density of the forecast. Various smoothing techniques exist, the most popular of them being kernel estimation and spline smoothing. These are discussed, for example, in Silverman (1986) or Härdel (1990, Ch 3).

- In practical work it is helpful to represent these densities by graphs. STAR model density forecasts may well be multimodal. If we cant to stress this property, we may graph the densities for each time period; see Matzner-Lober, Gannoun and De Gooijer (1998).

- Another idea is to compute and graph so-called highest-density regions. This is a compact way of representing these densities for a number of forecast periods simultaneously. Hyndman (1995) discusses the use of highest density regions in representing forecasts and provides an example of bimodal density forecasts.

- Let x be a continuous random variable with probability density $f(x)$.

- The highest density region $HDR_\alpha(x) = \{x: f(x) \geq f_\alpha\}$

where $f_\alpha > 0$ is such that the probability of a given x having a density that at least equal f_α is α.

- Any point x belonging to a highest density region has a higher density than any point outside this region.
- If the density is multi-modal, then a given highest-density region may consist of disjoint subsets of points depending on α.