Stationary Stochastic Time Series Models

Series de Tiempo

UTDT

January 2020

Introduction

- Uses of Univariate Times Series models:
  1. Forecasting: Need to find a model to characterize the series which then is used to produce a forecast.
  2. Finding which process an exogenous driving variable follows (e.g. A process for Dividends is usually assumed for calculating the fundamental value of Stock prices)

- Modeling a Time Series
- Consider a time series

$$(x_1, x_2 .... x_n)$$

as the realization of a stochastic process with distribution given by

$$p(x_1, x_2, ..., x_n).$$

- In general, fully specifying p is too ambitious: first and second moments

$$n \text{ means} : E(x_1), E(x_2), ..., E(x_n)$$
$$n \text{ variances} : V(x_1), V(x_2), ..., V(x_n)$$
$$\frac{n(n-1)}{2} \text{ covariances:} Cov(x_i, x_j), i < j.$$

- However, we only have n data points to estimate $2n + \frac{n(n-1)}{2}$ parameters!

Stationarity

Definition

A stochastic process is said to be **strictly stationary** *if its properties are unaffected by a change in the time origin, that is*

$$(\forall l)p(x_1, x_2, ..., x_n) = p(x_{1+l}, x_{2+l}, ..., x_{n+l}) \tag{1}$$

Definition

A stochastic process is said to be **weak stationary** *if the first and second moments exist and do not depend on time.*

$$E(x_1) = E(x_2) = ... = E(x_t) = \mu \tag{2}$$

$$V(x_1) = V(x_2) = ... = V(x_t) = \sigma^2 \tag{3}$$

$$Cov(x_t, x_{t-k}) = Cov(x_{t+l}, x_{t-k+l}) = \gamma_k \tag{4}$$

Comments

- Covariances are functions only of the lag k, and not of time. These are usually called **autocovariances**.
- From equations 3 and 4 we can obtain the **autocorrelations**

$$\rho_k = \frac{Cov(x_1, x_2)}{\sqrt[2]{V(x_1)V(x_2)}} = \frac{\gamma_k}{\sigma^2} = \frac{\gamma_k}{\gamma_o} \tag{5}$$

- The **autocorrelations** considered as a function of k are referred to as the autocorrelation function, **ACF**
- Notice:

$$\gamma_k = Cov(x_t, x_{t-k}) = Cov(x_{t-k}, x_t) = Cov(x_t, x_{t+k}) = \gamma_{-k} \tag{6}$$

Modelling a time Series using Moving Average and Autoregressive Processes: Using the ACF Function to identify a series.

- Lets define $y_t = (x_t - \mu)$ and consider the following first order M.A. Process, MA(1), for y_t:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Where ε_t is such that

$$\begin{aligned} E(\varepsilon_t) &= 0 && (7) \\ V(\varepsilon_t) &= \sigma^2 && (8) \\ Cov(\varepsilon_t, \varepsilon_{t-k}) &= 0 \text{ for all } k. && (9) \end{aligned}$$

- For this process we can:

1. Calculate the first and second moments of y_t and check the stationarity condition
2. Check its ACF in oder to use this function to Identify the process.

$$E(y_t) = 0$$

$$E(y_t)^2 = E(\varepsilon_t + \theta_1 \varepsilon_{t-1})^2 = \sigma^2(1 + \theta_1^2)$$

$$\begin{aligned} E(y_t y_{t-k}) &= E(\varepsilon_t + \theta_1 \varepsilon_{t-1})(\varepsilon_{t-k} + \theta_1 \varepsilon_{t-k-1}) \\ &\begin{cases} \sigma^2 \theta_1 & \text{for } k = 1 \\ 0 & \text{for } k > 1 \end{cases} \end{aligned}$$

$$\rho_k = \begin{cases} \theta_1 / (1 + \theta_1^2) & \text{for } k = 1 \\ 0 & \text{for } k > 1 \end{cases}$$

- Notice:

1. The MA(1) satisfies the Stationarity Conditions
2. The autocorrelation function, ρ_k, for a MA(1) has only the first term different from zero.

A Moving Average of order q: MA(q)

- A moving average process of order q can be written as:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}, \quad t = 1, ..., T$$

  We can calculate the first and second moments to check stationarity and also use the ACF function to Identify the series:

- $$E(y_t) = 0$$

$$E(y_t)^2 = \sigma^2 (1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)$$

$$E(y_t y_{t-k}) = \begin{cases} \sum_{j=0}^{q} \sigma^2 \theta_j \theta_{j+k} & \text{for } k = 1, 2, ..., q \\ 0 & \text{for } k > q \end{cases}$$

$$\rho_k = \begin{cases} \sum_{j=0}^{q} \sigma^2 \theta_j \theta_{j+k} / \sum_{j=0}^{q} \theta_j^2 & \text{for } k = 1, 2, ..., q \\ 0 & \text{for } k > q \end{cases}$$

- The ACF function of a MA(q) process has the first q terms different from zero and then all equal to zero.

Wold's Decomposition Theorem

Theorem

Every weakly stationary, purely non-deterministic, stochastic process $(x_t - \mu)$ can be written as a linear combination of uncorrelated random variables. The representation is given by:

$$\begin{aligned}(x_t - \mu) &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... \\ &= \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} \quad \textit{where } \theta_0 = 1\end{aligned} \tag{10}$$

Where ε_t is such that

$$E(\varepsilon_t) = 0 \tag{11}$$

$$V(\varepsilon_t) = \sigma^2 \tag{12}$$

$$Cov(\varepsilon_t, \varepsilon_{t-k}) = 0 \textit{ for all } k. \tag{13}$$

- NB, the MA(1) and the MA(q) are special cases of the theorem and therefore stationary.

Modelling a Time Series as an Autoregressive Model

An autoregressive process of order p, is written as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t, \quad t = 1, ..., T$$

This will be denoted y_t~AR(p)

- We will check

1. Under which conditions this processes are stationary
2. Whether we can use the ACF function to identify AR(p) processes.

- Consider the following first order autorregressive process, AR(1),

$$y_t = \phi_1 y_{t-1} + \varepsilon_t \qquad t = 1, ..., T.$$

Notice that if this relationship is valid for time t, it should also be valid for time $t-1$, that is

Example

$$y_{t-1} = \phi_1 y_{t-2} + \varepsilon_{t-1}$$

Substituting, we get:

$$\begin{aligned} y_t &= \phi_1 (\phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi_1^2 y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \end{aligned}$$

and Iterating:

$$y_t = \phi_1^j y_{t-j} + \phi_1^{j-1} \varepsilon_{t-(j-1)} + \phi_1^{j-2} \varepsilon_{t-(j-2)} + ... + \phi_1 \varepsilon_{t-1} + \varepsilon_t$$

- If $|\phi| < 1$, the deterministic component of y_t is negligible if j is large enough

$$y_t = \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j} \tag{14}$$

- An $AR(1)$ may be written as a $MA(\infty)$ in which the coefficient of ε_{t-j} is ϕ_1^j.

AR(1) stationary process

Lemma

When $|\phi| < 1$, then the AR (1) satisfies the conditions for stationarity

Proof.

The mean exists and does not depend on t

$$E(y_t) = E(\sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}) = 0$$

The Variance exists and does not depend on time

$$\begin{aligned} V(y_t) &= V(\sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}) \underset{E(y_t)=0}{=} E(\sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j})^2 \\ &= E(\sum_{j=0}^{\infty} \phi_1^{2j} \varepsilon_{t-j}^2) = \sigma^2 \sum_{j=0}^{\infty} \phi_1^{2j} = \frac{\sigma^2}{(1-\phi_1^2)} \end{aligned}$$

Proof.

(continues)
The Autocovariances exist and only depend on lag k

$$\begin{aligned} E(y_t y_{t-k}) &= \phi_1 E(y_{t-1} y_{t-k}) + E(\varepsilon_t y_{t-k}) \\ \gamma_k &= \phi_1 \gamma_{k-1} + E(\varepsilon_t y_{t-k}) \end{aligned}$$

Notice that:

$$E(\varepsilon_t y_{t-k}) = E[\varepsilon_t(\phi_1^{j-1}\varepsilon_{t-k-(j-1)} + ... + \phi_1 \varepsilon_{t-(k-1)} + \varepsilon_{t-k})]$$

Given ε_t is white noise:

$$\gamma_k = \phi_1 \gamma_{k-1}$$

Thus

$$\rho_k = \phi_1 \rho_{k-1} \tag{15}$$

□

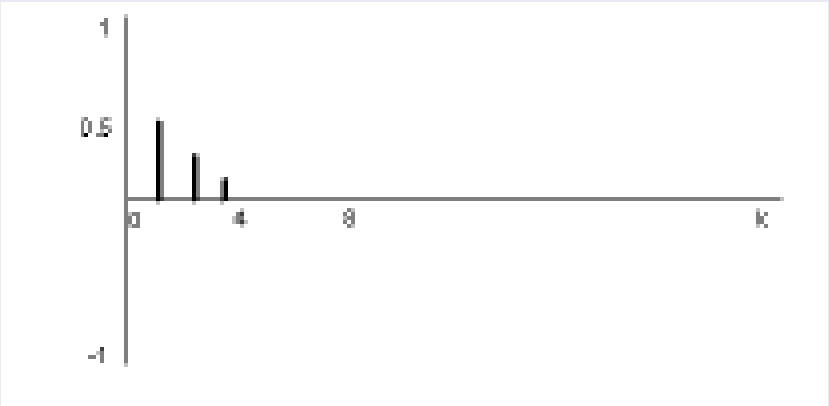
Comments

- Whenever the process is stationary the autocorrelation function declines exponentially:

$$\rho_k = \phi_1^k \rho_0.$$

Example

For $\phi = 0.5$:



General conditions for stationarity: The use of the lag operator

Definition

The lag operator, L, is defined by the transformation

$$Ly_t = y_{t-1}$$

- Notice that the lag operator may also be applied to y_{t-1} yielding

$$Ly_{t-1} = y_{t-2}$$

- Substitution, yields:

$$L^k y_t = y_{t-k} \qquad \text{for } k \geq 0$$

- The lag operator can be manipulated in a similar way to any algebraic quantity.

The lag operator

Example

Let us reproduce the $MA(\infty)$ *representation of the* $AR(1)$ *process:*

$$y_t = \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}$$

where we assume $|\phi_1| < 1$ *. Using* L *this may be written as:*

$$y_t = \sum_{j=0}^{\infty} (\phi_1 L)^j \varepsilon_t = \varepsilon_t / (1 - \phi_1 L)$$

This can be rearranged in the following way

$$(1 - \phi_1 L) y_t = \varepsilon_t$$

$$y_t = \phi_1 y_{t-1} + \varepsilon_t$$

Autoregressive processes using Lag Operators

An AR(p) process may be written as,

$$(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p) y_t = \varepsilon_t, \qquad t = 1, ..., T$$

or

$$\phi(L) y_t = \varepsilon_t, \qquad t = 1, ..., T$$

where $\phi(L) = (1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)$.

Stationarity Conditions for Autoregressive processes

- The Stationarity condition may be expressed in terms of the roots of the polynomial of order p in L. Consider an $AR(1)$

$$(1 - \phi_1 L) y_t = \varepsilon_t, \qquad t = 1, ..., T.$$

When we consider the root of $(1 - \phi_1 L) = 0$, that is $L = 1/\phi_1$, we see it is greater than 1 (in absolute value) whenever $|\phi_1| < 1$.Then the process seems to be stationay whenever $L > 1$.

- In general: an $AR(p)$ is said to be stationary when all the roots of the polynomial $(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)$ lie outside the unit circle.

Moving Average processes using Lag Operators

Consider the following MA(q) process

$$y_t = (1 + \theta_1 L + \theta_2 L^2 + ... + \phi_q L^q)\varepsilon_t \qquad t = 1, ..., T$$

or

$$y_t = \theta(L)\varepsilon_t$$

where $\theta(L) = (1 + \theta_1 L + \theta_2 L^2 + ... + \phi_q L^q)$.

Remark A $MA(q)$ process is said to be **invertible** if all the roots of the polynomial $(1 + \theta_1 L + \theta_2 L^2 + ... + \phi_q L^q)$ lie outside the unit circle.

Autoregressive Moving Average Process

Definition

An autoregressive moving average process of order (p, q), denoted as ARMA(p, q) is written as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}$$

or

$$\Phi(L) y_t = \Theta(L) \varepsilon_t$$

Autoregressive Moving Average Process

- AR(p) and MA(q) are special cases of the ARMA(p, q) process.
- The stationarity of an ARMA process depends solely on its AR part
  - ARMA is stationary if $\Phi(L) = 0$ lies outside the unit circle
- Its invertibility depends only on its MA part
  - ARMA is invertible if $\Theta(L) = 0$ lies outside the unit circle
- If both conditions hold, ARMA can be written as $AR(\infty)$ or $MA(\infty)$

Example

ARMA (1,1)

$$y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Autocovariance function

$$\begin{aligned} \gamma_k &= E(y_t y_{t-k}) = \phi_1 E(y_{t-1} y_{t-k}) + E(\varepsilon_t y_{t-k}) + \theta_1 E(\varepsilon_{t-1} y_{t-k}) \\ &= \phi_1 \gamma_{k-1} + E(\varepsilon_t y_{t-k}) + \theta_1 E(\varepsilon_{t-1} y_{t-k}) \end{aligned}$$

Example

For $k = 0$

$$\gamma_0 = \phi_1 \gamma_1 + E(\varepsilon_t(\phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1})) + \theta_1 E(\varepsilon_{t-1}(\phi_1 y_{t-1} + \varepsilon_t + \phi_1 \varepsilon_{t-1})$$

where

$$E(\varepsilon_t(\phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1})) = \sigma^2$$

$$E(\varepsilon_{t-1}(\phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2}) + \varepsilon_t + \theta_1 \varepsilon_{t-1})) = (\phi_1 + \theta_1)\sigma^2$$

Then:

$$\gamma_0 = \phi_1 \gamma_1 + \sigma^2 + \theta_1(\phi_1 + \theta_1)\sigma^2$$

Example

When $k = 1$

$$\gamma_1 = \phi_1 \gamma_0 + \theta_1 E(\varepsilon_{t-1}(\phi_1 y_{t-2} + \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2}))$$

Thus,

$$\gamma_1 = \phi_1 \gamma_0 + \theta_1 \sigma^2$$

For $k \geq 2$

$$\gamma_k = \phi_1 \gamma_{k-1}$$

Partial Autocorrelations

- The ACF can be used to check whether a process is a MA(q), and to determine q.
- The ACF for an AR declines exponentially
- However, we cannot guess the order of the AR from the plot of the ACF
- Thus, we need to use the Partial Autocorrelation Function, PACF, to determine the order of the autoregressive process.

Partial Autocorrelation

- Consider an AR(1) process, the correlation between y_t and y_{t-2} comes through the correlation each other has with y_{t-1}
- The k^{th} partial autocorrelation, $\phi_k = \phi_{kk}$, function measures the correlation which comes only from the direct effect of the k^{th} lag.
- To find how many partial autocorrelations are different from zero, we use the Yule Walker equations

Yule Walker equations

For an autoregressive process of order p the Yule-Walker equations are given by the following recursion formulae;

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \phi_3 \rho_{k-3} + ... + \phi_p \rho_{k-p} \quad \textit{for } k = 1, ..., p.$$

Then we just need to set $k = p$ or $\phi_p = \phi_{kk}$ and solve the following system of equations:

$$\rho_k = \phi_{11} \rho_{k-1} + \phi_{22} \rho_{k-2} + \phi_{33} \rho_{k-3} + ... + \phi_{kk}.$$

Yule Walker equations

Giving values $k = 1....k$:

$$\begin{aligned}
\rho_1 &= \phi_{11}\rho_0 + \phi_{22}\rho_1 + \phi_{33}\rho_2 + ... + \phi_{kk}\rho_{k-1} \quad \text{for } k = 1 \\
\rho_2 &= \phi_{11}\rho_1 + \phi_{22}\rho_0 + \phi_{33}\rho_1 + ... + \phi_{kk}\rho_{k-2} \quad \text{for } k = 2 \\
&\ldots \\
\rho_k &= \phi_{11}\rho_{k-1} + \phi_{22}\rho_{k-2} + \phi_{33}\rho_{k-3} + ... + \phi_{kk}\rho_0 \quad \text{for } k = k
\end{aligned}$$

- It is a k by k system which can be solved for $\phi_{ii}(\rho_1, ..., \rho_k)$ for $i = 1, 2, ...k$.

Yule Walker equations

- The PACF of an AR(p) has the first p terms different from zero and the rest equal to zero. The empirical methodology consists in finding which ϕ_{kk} are not significantly different from zero.

Example

AR (1)

$$\rho_k = \phi_1 \rho_{k-1}$$

Then,

$$\rho_k = \phi_{11} \rho_{k-1} \qquad \text{since } p = k = 1,$$

$$\begin{aligned} \rho_1 &= \phi_{11} \rho_0 \qquad \text{for } k = 1. \\ &\text{or} \\ \rho_1 &= \phi_{11} \end{aligned}$$

Example

AR (2)

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$$

Thus,

$$\rho_k = \phi_{11} \rho_{k-1} + \phi_{22} \rho_{k-2} \text{ since } p = k = 2,$$

Giving values to k

$$\begin{aligned} \rho_1 &= \phi_{11} + \phi_{22}\rho_1 && \text{for } k = 1, \\ \rho_2 &= \phi_{11}\rho_1 + \phi_{22} && \text{for } k = 2, \end{aligned}$$

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}.$$

Identifying the series through ACF/PACF

- *For an AR(p) Process:* The ACF declines exponentially and the PACF is zero for lags greater than p
- *For a MA(p) Process:* The ACF is zero for lags greater than q and the PACF declines exponentially

Box Jenkins Methodology:

Using *sample* information, we might calculate sample ACF and PACF to try to identify the right model.

1. Transform the data, if necessary, so that the assumption of covariance stationarity is a reasonable one.
2. Make an initial guess of small values of p and q for an ARMA(p, q) model that might describe the transformed series.
3. Estimate the parameters in $\phi(L)$ and $\theta(L)$
4. Perform diagnostic analysis to confirm that the model is indeed consistent with the observed features of the data.

What's next:

- Empirical properties of sample analogs
- Estimation

Correlogram

- Basic tool
- An inspection of the correlogram may lead to the conclusion that the series is random, or that exhibits a pattern of serial correlation that which perhaps can be modeled by a particular stochastic process

Sample ACF and PACF

It can be proved that asymptotically:

$$\begin{aligned} \widehat{\rho_i} &\approx N(0, \frac{1}{T}) \text{ under } H_0 : \rho_i = 0 \\ \widehat{\phi_{ii}} &\approx N(0, \frac{1}{T}) \text{ under } H_0 : \phi_{ii} = 0 \end{aligned}$$

where T is the sample size

Significance test

- In order to identify using ACF we should test whether the different parameters ρ_k are different from zero.
- Box Pierce Q- Statistic

$$Q = T \sum_{i=1}^{k} \widehat{\rho}_i^{2} \tilde{} \chi^2(k) \text{ under } H_0 : \rho_1 = ... = \rho_k = 0$$

If I don't reject H_0, then, $\{\varepsilon_t\}$ is a white noise. This test has low power, even in large samples.

- Corrected version

$$Q^* = T(T+2)\sum_{i=1}^{k}\frac{\widehat{\rho}_i^{\,2}}{T-i} \tilde{} \chi^2(k) \text{ under } H_0 : \rho_1 = ... = \rho_k = 0$$

- When the k autocorrelations are calculated for an ARMA(p,q) model, we lose degrees of freedom. Thus

$$Q^* \tilde{} \chi^2(k-p-q)$$

Significance test

- When we use identifying tools such as :

$$\begin{aligned} \widehat{\rho_i} &\approx N(0, \frac{1}{T}) \text{ under } H_0 : \rho_i = 0 \\ \widehat{\phi_{ii}} &\approx N(0, \frac{1}{T}) \text{ under } H_0 : \phi_{ii} = 0 \end{aligned}$$

  we find that these tools won't tell us neither whether the preferred model is mispecified, nor what to do when two different models seem to be equally valid.
- We will need to estimate these models.

Maximum Likelihood Estimation

- Usually when we estimate ARMA(p, q) models we evaluate the conditional maximum likelihood.
  - Assume that the first $max(p, q)$observations are known.
- Asymptotically, CMLE is equivalent to the MLE and is easier to calculate

Example

AR(1)

$$f(y_T, y_{T-1}, ..., y_2 | y_1, \phi_1, \sigma^2) = \prod_{i=2}^{T} f(y_i | y_{i-1}, \phi_1, \sigma^2)$$

where

$$f(y_i | y_{i-1}, \phi_1, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_t - \phi_1 y_{t-1})^2}{2\sigma^2}}$$

The objective is to maximize:

$$\mathcal{L} = -(T-1) log(2\pi)^{.5} - (T-1) \log \sigma - \frac{\sum_{t=2}^{T} (y_t - \phi_1 y_{t-1})^2}{2\sigma^2}$$

Model Selection Criteria

- Akaike Criteria (AIC):

$$AIC(p,q) = \log \hat{\sigma}^2 + 2(p+q)\,T^{-1}$$

- Shwartz Criteria (SC):

$$SC(p,q) = \log \hat{\sigma}^2 + (p+q)\,T^{-1} \log(T)$$

- The idea is to weigh likelihood and number of lags.
- It is best when we compare models with the same number of lags
- Overparametrized models have less forecast ability

- We choose the model with the lower AIC/SC
- SC picks more parsimonious models
- SC has better large sample properties, while AIC picks overparametrized models when $T \longrightarrow \infty$
- For small samples, AIC may outperform the others
- If they pick different models, as SC picks the most parsimonious you should check the residuals and with AIC you should check for the significance of the parameters

Forecast

- In practice, we only have an approximation to the "true GDP"
  - identification and estimation errors

Minimum Mean Square Error Forecast

We start with conditional mean

- Serves as benchmark
- It generates the forecast with the minimum mean square error
  ie, if the model is correct, there is no other statistic that produces smaller forecast errors

Given observations up to, and including y_T :

$$\widehat{y}_{T+l|T} = E(y_{T+l}|I_T)$$

where I_T is the information set up to T

Notice that for any other forecast $\widetilde{y}_{T+l|T}$, the forecast error maybe divided in two:

$$y_{T+l} - \widetilde{y}_{T+l|T} = [y_{T+l} - \widehat{y}_{T+l|T}] + [\widehat{y}_{T+l|T} - \widetilde{y}_{T+l|T}]$$

Squaring terms and conditional on T:

$$MSE(\widetilde{y}_{T+l|T}) = Var(y_{T+l}) + [\widetilde{y}_{T+l|T} - E(y_{t+l}|I_T)]^2$$

One step ahead forecast

Consider the following stationary and invertible ARMA (p, q):

$$\begin{aligned} y_{T+1} = \quad & \phi_1 y_T + \phi_2 y_{T-1} + ... + \phi_p y_{T-p+1} + \\ & \varepsilon_{T+1} + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1} + ... + \theta_q \varepsilon_{T-q+1} \end{aligned}$$

Then,

$$\begin{aligned} \widehat{y}_{t+1|T} = \quad & \phi_1 y_T + \phi_2 y_{T-1} + ... + \phi_p y_{T-p+1} + \\ & \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1} + ... + \theta_q \varepsilon_{T-q+1} \end{aligned}$$

We need values for $\{\varepsilon_t\}$ to obtain numerical values for the forecast

Example

AR(1)

$$y_{T+l} = \phi_1 y_{T+l-1} + \varepsilon_{T+l} \qquad \text{at time } T+l$$

$$\widehat{y}_{T+l|T} = \phi_1 \widehat{y}_{T+l-1|T} \qquad l = 1, 2, ...$$

The initial value is $\widehat{y}_{T|T} = y_T$. *Then:*

$$\widehat{y}_{T+l|T} = \phi_1^l y_T$$

Example

(continues)
Let's calculate the forecasting error:

$$\begin{aligned} y_{T+l} - \widehat{y}_{T+l|T} &= \phi_1 y_{T+l-1} + \varepsilon_{T+l} - \phi_1^l y_T \\ &= \phi_1^l y_T + \varepsilon_{T+l} + \phi_1 \varepsilon_{T+l-1} + \\ &\quad \phi_1^2 \varepsilon_{T+l-2} + ... + \phi_1^{l-1} \varepsilon_{T+l} - \phi_1^l y_T \end{aligned}$$

The forecasting error variance is given by:

$$\begin{aligned} V(y_{T+l} - \widehat{y}_{T+l|T}) &= V(\varepsilon_{T+l} + \phi_1 \varepsilon_{T+l-1} + \phi_1^2 \varepsilon_{T+l-2} + ... + \phi_1^{l-1} \varepsilon_{T+l}) \\ &= (1 + \phi_1^2 + \phi_1^4 + ... + \phi_1^{2(l-1)}) \sigma^2 \end{aligned}$$

Example

MA(1)
On $T+1$, the equation for MA(1) is given by:

$$y_{T+1} = \varepsilon_{T+1} + \theta_1 \varepsilon_T$$

Thus, in general,

$$\widehat{y}_{T+l|T} = \hat{\varepsilon}_{T+l|T} + \theta_1 \hat{\varepsilon}_{T+l-1|T}$$

$$\begin{aligned} \widehat{y}_{T+l|T} &= \theta_1 \varepsilon_T \qquad \text{for } l = 1 \\ &= 0 \qquad \text{for } l > 1. \end{aligned}$$

Example

(*continues*)
The variance of the forecast error for a MA(1) is

$$\begin{aligned} V(y_{T+l} - \widehat{y}_{T+l|T}) &= \sigma^2 && \text{for } l = 1 \\ &= (1 + \theta_1^2)\sigma^2 && \text{for } l > 1 \end{aligned}$$

Example

$ARMA(1,1)$

$$y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

$$\begin{aligned} \widehat{y}_{T+l|T} &= \phi_1 y_T + \theta_1 \varepsilon_T && \text{for } l = 1 \\ &= \phi_1 \widehat{y}_{T+l-1|T} && \text{for } l > 1 \\ &= \phi_1^l y_T + \phi_1^{l-1} \theta_1 \varepsilon_T \end{aligned}$$

Measuring the Accuracy of Forecasts

Root Mean Squared Error:

$$RMSE = \sqrt{\frac{1}{l} \sum_{i=T+1}^{T+l} (\widehat{Y}_{i|T} - Y_i)^2}$$

Mean Absolute Error

$$MAE = \frac{1}{l} \sum_{i=T+1}^{T+l} |\widehat{Y}_{i|T} - Y_i|$$

Remark Which indicator should be used depends of the purpose of the forecasting exercise. The RMSE will penalize big errors more than the MAE measure.

Vector Autoregressions

Series de Tiempo

Birkbeck

January 2022

Uses of Vector Autoregressions:

- Forecast
- Testing Linear Rational Expectations Models.
- Granger Causality.
- Impulse Response Analysis.
- Variance Decomposition.

STRUCTURE OF THE LECTURE

- Definition and Properties.
- Estimation
- VAR analysis.

VAR

Definition

A vector autoregressive (VAR) is simply an autoregressive process for a vector of variables.

Let us define $W_t = \begin{bmatrix} x_t \\ y_t \end{bmatrix}$, *a matrix* A_{2X2} *and* $\varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$.

where

$$E(\varepsilon_t) = 0, \qquad E(\varepsilon_t \varepsilon_s') = \begin{cases} \Omega & t = s \ (\Omega = \Omega', \ c'\Omega c > 0, \ c \neq 0), \\ 0 & \text{otherwise} \end{cases}$$

VAR (1)

- Then a **VAR(1)** may be written as

$$W_t = AW_{t-1} + \varepsilon_t$$

or

$$x_t = a_{11}x_{t-1} + a_{12}y_{t-1} + \varepsilon_{1t},$$

$$y_t = a_{21}x_{t-1} + a_{22}y_{t-1} + \varepsilon_{2t},$$

VAR(p)

- A VAR of order p can be written as

$$W_t = A_1 W_{t-1} + A_2 W_{t-2} + ... + A_p W_{t-p} + \varepsilon_t$$

  Using the lag operator

$$(I - A_1 L - A_2 L^2 - ... - A_p L^p) W_t = \varepsilon_t$$

- The VAR is covariance stationary if all the values of L satisfying $|I - A_1 L - A_2 L^2 - ... - A_p L^p| = 0$ lie outside the unit circle.

The Autocovariance Matrix

- For a covariance stationary n dimensional vector process we may define the *autocovariance function* for a VAR in a way similar to the univariate case:

$$\Gamma_{k\ (nxn)} = E(W_t W'_{t-k}) \quad \textit{where} \quad \Gamma_{k(ij)} = cov(W_{i,t}, W_{j,t-k})$$

Example

Using the above two variables VAR we get the following:

$$\Gamma_{k\ (nxn)} = E(W_t W'_{t-k}) = \begin{bmatrix} E(x_t x_{t-k}) & E(x_t y_{t-k}) \\ E(y_t x_{t-k}) & E(y_t y_{t-k}) \end{bmatrix}$$

Comments

- Contrary to the univariate case $\Gamma_k \neq \Gamma_{-k}$, instead the correct relationship is:

$$\Gamma_k' = \Gamma_{-k}$$

Proof.

Leading $E(W_t W_{t-k}')$ k periods we get $\Gamma_{k\ (\text{nxn})} = E(W_{t+k} W_t')$. Then, transposing, we obtain:

$$\Gamma_{k\ (\text{nxn})}' = E(W_t W_{t+k}') = \Gamma_{-k}$$

□

Intuition There is no reason why $E(x_t y_{t-1})$ should be equal to $E(x_{t-1} y_t)$.

Example

VAR (1) *autocovariance function*

$$\Gamma_{k\ (nxn)} = E(W_t W'_{t-k}) = AE(W_{t-1} W'_{t-k}) + E(\varepsilon_t W'_{t-k})$$

Thus, for $k \geq 1$ *:*

$$\Gamma_{k\ (nxn)} = A\Gamma_{k-1}.$$

Example

(continues) For $k = 0$*:*

$$E(W_t W_t') = AE(W_{t-1} W_{t-1}')A' + E(\varepsilon_t \varepsilon_t')$$

or

$$\Gamma_0 = A\Gamma_0 A' + \Omega$$

In order to obtain Γ_0 *we use the vec operator*

$$vec(\Gamma_0) = vec(A\Gamma_0 A') + vec(\Omega) = (A \otimes A)vec(\Gamma_0) + vec(\Omega),$$

Using that $vec(ABC) = (C' \otimes A)vec(B)$ *:*

$$vec(\Gamma_0) = (I_{(n)^2} - (A \otimes A))^{-1} vec(\Omega).$$

The companion form

- Notice that a VAR(p) may always be re-written as a VAR(1) by defining a vector H_t such that:

$$H_t = FH_{t-1} + \nu_t$$

where

$$H_t = \begin{bmatrix} x_t \\ y_t \\ . \\ x_{t-i} \\ y_{t-i} \\ . \\ x_{t-(p-1)} \\ y_{t-(p-1)} \end{bmatrix} \quad F = \left[\begin{array}{cc|cc} A_1 & A_2 & ... & A_p \\ I_{2\times 2} & 0 & ... & 0 \\ ... & ... & I & 0 \end{array}\right] \quad \nu_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ .. \\ 0 \end{bmatrix}$$

The companion form

- Then, the VAR in the Companion form can be expressed in the following way

$$H_t = FH_{t-1} + \nu_t \qquad E(\nu_t \nu_t') = \begin{cases} Q & t = s \\ 0 & \text{otherwise} \end{cases}$$

where

$$Q_{(\text{np} \times \text{np})} = \begin{bmatrix} \Omega & 0 & 0 & \dots & 0 \\ 0 & . & & & \\ . & . & & & \\ 0 & 0 & & \dots & 0 \end{bmatrix}$$

- The variance covariance matrix can be found noticing that

$$E(H_t H_t') = FE(H_{t-1} H_{t-1}')F' + Q$$

or

$$\Sigma = F\Sigma F' + Q \qquad \text{where } \Sigma = E(H_t H_t').$$

$$\Sigma = \begin{bmatrix} \Gamma_0 & \Gamma_1 & \dots & \Gamma_{p-1} \\ \Gamma_1' & \Gamma_0 & & \Gamma_{p-2} \\ & & & \\ \Gamma_{p-1}' & \Gamma_{p-2}' & & \Gamma_0 \end{bmatrix}$$

with Γ_p the autocovariance of the original process

- If the process is covariance stationary, then the unconditional variance can be calculated simply using vec operators, i.e.,

$$vec(\Sigma) = vec(F\Sigma F') + vec(Q) = (F \otimes F)vec(\Sigma) + vec(Q),$$

Then the unconditional variance can be obtained as

$$vec(\Sigma) = (I_{(np)^2} - (F \otimes F))^{-1} vec(Q).$$

- Notice as well that the j^{th} autocovariance function of H (denoted Σ_j) can be found by post-multiplying by H'_{t-j} and taking expectations.

$$E(H_t H'_{t-j}) = FE(H_{t-1} H'_{t-j}) + E(\nu_t H'_{t-j})$$

- Thus,

$$\Sigma_k = F\Sigma_{k-1} \quad para\ k = 1, 2, ...$$

or

$$\Sigma_k = F^k\Sigma.$$

- The k^{th} autocovariance Γ_k of the original process W_t is given by the n first rows and n columns of $\Sigma_k = F\Sigma_{k-1}$:

$$\Gamma_k = A_1\Gamma_{k-1} + A_2\Gamma_{k-2} + ... + A_p\Gamma_{k-p} \qquad k = p, p+1, p+2, ...$$

The Conditional likelihood for a vector autoregression.

- Let W_t denote an $(n_{\times}1)$ vector which we assume follows a p^{th} order Gaussian VAR.

$$W_t = A_1 W_{t-1} + A_2 W_{t-2} + ... + A_p W_{t-p} + \varepsilon_t \qquad \varepsilon_t \tilde{} N(0, \Omega),$$

- The approach is to condition on the first p observations $(W_0, ...W_{-p+1})$ and to base the estimation on the last T observations $(W_T, ...W_1)$.

$$f(W_T, W_{T-1}, W_{T-2}....W_1 | W_0, ., W_{-p+1}; \Theta)$$

and maximize with respect to Θ, where Θ is a vector that contains the elements of $A_1, A_2, A_3, ...A_p$ and Ω.

- Then,

$$W_t | W_{t-1}, ..., W_{-p+1} \tilde{} N\left(A_1 W_{t-1} + ... + A_p W_{t-p}, \Omega_{n \times n}\right)$$

- It will be convenient to stack the p lags in a vector x_t,.

$$x_t = \begin{bmatrix} \underbrace{W_{t-1}}_{n \times 1} \\ \underbrace{W_{t-2}}_{n \times 1} \\ . \\ \underbrace{W_{t-p}}_{n \times 1} \end{bmatrix}_{np \times 1}$$

- let Π' denote the following $n_{\times} np$ matrix :

$$\Pi' = [A_1, A_2, A_3, ...A_p]_{n \times np}$$

- Thus, the conditional mean is just

$$\Pi' x_t$$

- Therefore,

$$W_t | W_{t-1}, ..., W_{-p+1} \tilde{} N(\Pi' x_t, \Omega)$$

or

$$\begin{aligned} & f(W_t | W_{t-1}, .., W_{-p+1}; \Theta) \\ = \quad & (2\pi)^{-n/2} |\Omega^{-1}|^{.5} \exp[(-1/2)(W_t - \Pi' x_t)' \Omega^{-1} (W_t - \Pi' x_t)] \end{aligned}$$

- The joint density conditional on the first p observations can be written as:

$$\begin{aligned} & f(W_T, W_{T-1}, W_{T-2} .... W_1 | W_0, ., W_{-p+1}; \Theta) \\ = & \prod_{t=1}^{T} f(W_t | W_{t-1}, .., W_{-p+1}; \Theta) \end{aligned}$$

- Taking logs

$$\begin{aligned} L(\Theta) &= \sum_{t=1}^{T} \ln[f(W_t | W_{t-1}, .., W_{-p+1}; \Theta)] \\ &= -(Tn/2) log(2\pi) + (T/2) log|\Omega^{-1}| \\ &\quad -(1/2) \sum_{t=1}^{T} (W_t - \Pi' x_t)' \Omega^{-1} (W_t - \Pi' x_t) \end{aligned}$$

Estimating the parameters in A

It turns out to be that the maximum likelihood estimator is

$$\hat{\Pi}' = [\sum_{t=1}^{T} W_t x_t'][\sum_{t=1}^{T} x_t x_t']^{-1}$$

where the j column is just

$$\hat{\pi}_{j\ (1\times \text{np})} = [\sum_{t=1}^{T} W_{jt} x_t'][\sum_{t=1}^{T} x_t x_t']^{-1}$$

The Maximum likelihood estimator of

Ω

- We can now "concentrate" the likelihood using the previous results to find the MLE estimator of Ω
- Evaluate the likelihood at the estimate of Π

$$L(\Omega, \hat{\Pi}) = -(Tn/2)log(2\pi) + (T/2)log|\Omega^{-1}| - (1/2)\sum_{t=1}^{T} \hat{\varepsilon}_t' \Omega^{-1} \hat{\varepsilon}_t$$

- Using the differentation rules $\frac{\partial x'Ax}{\partial A} = xx'$, $\frac{\partial \log |A|}{\partial A} = \left(A^{-1}\right)'$ and taking the derivative of $L(\Omega, \hat{\Pi})$ with respect to Ω^{-1}

$$\frac{\partial L(\Omega, \hat{\Pi})}{\partial \Omega^{-1}} = (T/2)\Omega' - (1/2)\sum_{t=1}^{T} (\hat{\varepsilon}_t \hat{\varepsilon}_t').$$

- Equating this expression to zero we obtain the MLE of the variance-covariance matrix.

$$\hat{\Omega}' = (1/T)\sum_{t=1}^{T} (\hat{\varepsilon}_t \hat{\varepsilon}_t')$$

Comments

Row i, column i of $\hat{\Omega}$ is given by

$$\hat{\sigma}_i^2 = (1/T) \sum_{t=1}^{T} (\hat{\varepsilon}_{it}^2)$$

which is just the average squared residual from a regression of a variable of the VAR on the p lags of all variables
Therefore I can use OLS results to construct both $\hat{\Omega}$ and $\hat{\Pi}$.

Choosing the order of VAR

- The validity of the tests we carry out depend on having identified the order of the VAR correctly
- A simple way to do so is comparing likelihood ratios
  - These can be easily computed because we are using OLS
- Consider the likelihood function at is Maximum value of a VAR with p_0 lags:

$$L_0(\hat{\Omega}, \hat{\Pi}) = -(Tn/2)log(2\pi) + (T/2)log|\hat{\Omega}_0^{-1}| - (1/2)\sum_{t=1}^{T} \hat{\varepsilon}_t' \hat{\Omega}_0^{-1} \hat{\varepsilon}_t.$$

Consider now the last term of this equation,

$$(1/2)\sum_{t=1}^{T}\hat{\varepsilon}_t'\hat{\Omega}_0^{-1}\hat{\varepsilon}_t \underbrace{=}_{(\ a\ scalar)} TR((1/2)\sum_{t=1}^{T}\hat{\varepsilon}_t'\hat{\Omega}_0^{-1}\hat{\varepsilon}_t)$$

$$\underbrace{=}_{TR(A.B)=TR(B.A)} (1/2)TR(\sum_{t=1}^{T}\hat{\Omega}_0^{-1}\hat{\varepsilon}_t\hat{\varepsilon}_t')$$

$$\underbrace{=}_{\hat{\Omega}_0=\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t'/T)} (1/2)TR(\hat{\Omega}_0^{-1}T\hat{\Omega}_0)$$

$$= \quad (T/2)TR(I) = (nT)/2.$$

Thus,

$$L_0(\hat{\Omega},\hat{\Pi}) = -(Tn/2)log(2\pi) + (T/2)log|\hat{\Omega}_0^{-1}| - (nT)/2$$

- If we want to test the Hypothesis that the VAR has p lags against p_0 lags we calculate the likelihood for the VAR with p_1 lags ($p_1 > p_0$)

$$L_1(\hat{\Omega}, \hat{\Pi}) = -(Tn/2)log(2\pi) + (T/2)log|\hat{\Omega}_1^{-1}| - (nT)/2$$

and compute the likelihood ratio which is

$$\begin{aligned} & 2(L_1(\hat{\Omega}, \hat{\Pi}) - L_0(\hat{\Omega}, \hat{\Pi})) \\ = \quad & T(\log|\hat{\Omega}_1^{-1}| - \log|\hat{\Omega}_0^{-1}|) \tilde{} \chi^2(n^2(p_1 - p_0)) \\ & \text{under } H_0 \end{aligned}$$

- Sims (1980) proposed the following for small samples:

$$(T-k)(\log|\hat{\Omega}_1^{-1}| - \log|\hat{\Omega}_0^{-1}|)$$

where $k = np_1 = \max\{\text{number of parameters estimated per equation}\}$

Goodness of Fit Criteria

- Measures how good a model is relative to others
- Balance between fit and complexity
- Typically, we would like to minimize:

$$C(p) = -2max(logL) + \beta(number\ of\ freely\ estimated\ parameters)$$

- For Gaussian models, the maximized log-likelihood is proportional to

$$-(T/2)log|\Omega| \qquad\qquad (since|\Omega^{-1}| = 1/|\Omega|)$$

Goodness of Fit Criteria

- Hence, we choose p to minimize:

$$C(p) = Tlog|\Omega| + \beta(n^2 p)$$

- For example

AIC	$\beta = 2$ (Akaike information criterion)
SBC	$\beta = log(T)$ (Scharz Bayesian criterion)
HQ	$\beta = 2log(log(T))$ (Hannan-Quin criterion)

- Alternatively the Akaike's prediction error (FPE) criterion chooses p so that to minimize the expected one -step ahead squared forecast error:

$$FPE = [\frac{T + np + 1}{T - np - 1}]^n |\Omega|$$

Asymptotic Distribution of the VAR estimators

- MLE will give consistent estimators of Π and Ω
- Standard errors of $\hat{\Pi}$ are given by standard OLS formulas
- Let $\hat{\pi}_T = vec(\hat{\Pi}_T)$ denote the $nk_\times 1$ vector of coefficients resulting from OLS. Then

$$\sqrt{T}(\hat{\pi}_T - \pi) \xrightarrow{L} N(0, \Omega \otimes Q^{-1})$$

  where $Q = E(x_t x_t')$.
- Standard OLS t and F statistics applied to the coefficients of any single equation in the VAR are asymptotically valid.

Testing Rational Expectations Hypothesis

- These models usually impose non-linear cross equation restrictions between the parameters of the model which are tested using a likelihood ratio test
- Consider a first order bivariate VAR:

$$\begin{aligned} x_t &= a_{11}x_{t-1} + a_{12}y_{t-1} + \varepsilon_t \\ y_t &= a_{21}x_{t-1} + a_{22}y_{t-1} + \nu_t \end{aligned}$$

  where x_t is the interest rates differential and y_t is the first difference of the logs of the spot exchange rate
- Then uncovered interest parity can be written as

$$x_t = E_t y_{t+1}.$$

Testing Rational Expectations Hypothesis

- Condition on both sides of the previous equation on $t-1$,we get the following restrictions

$$\begin{bmatrix} 1 & 0 \end{bmatrix} A = \begin{bmatrix} 0 & 1 \end{bmatrix} A^2$$

which can be expressed as:

$$a_{11} = a_{22}a_{21}/(1-a_{21}) \qquad a_{12} = a_{22}^2/(1-a_{21})$$

- Estimate the unrestricted and the restricted model and perform a likelihood ratio test:

$$2(L_u - L_r) \tilde{} asymptotically \chi^2$$

Granger Causality

Definition

y fails to Granger-cause x if for all $s > 0$ the mean squared error of a forecast of x_{t+s} based on $(x_t, x_{t-1}, ...)$ is the same as the MSE of a forecast of x_{t+s} based on $(x_t, x_{t-1}, ...)$ and $(y_t, y_{t-1}, ...)$. For linear functions

$$MSE[E(x_{t+s}|x_t, x_{t-1}, ...)] = MSE[E(x_{t+s}|x_t, x_{t-1}, ..., y_t, y_{t-1}, ...)]$$

Remark Granger's reason for proposing this definition was that if an event Y is the cause of another event X, then the event Y should precede the event X.

Granger Causality

- The null hypothesis is that y fails to Granger - cause x
- We just regress both the general model

$$x_t = a_{11}x_{t-1} + a_{12}y_{t-1} + \varepsilon_{1t}$$

and the restricted model

$$x_t = a_{11}x_{t-1} + \varepsilon'_{1t}$$

and compare the residuals sum squares
$T(RRS(\varepsilon') - RRS(\varepsilon))/RRS(\varepsilon) \tilde{} \chi^2(1)$ (asymptotically)

Conditioning on the correct information set

- Omitting a relevant variable in the information set may give spurious results, since a variable that is thought to be useful for forecasting others, may be not longer useful once you condition on the right information.
- The question of whether a scalar y can help forecast another scalar x needs to be accommodated considering the information about z.
- Then, y fails to Granger-cause x if for all $s > 0$ the mean squared error of a forecast of x_{t+s} based on $(x_t, x_{t-1}, .z_t, z_{t-1}...)$ is the same as the MSE of a forecast of x_{t+s} based on $(x_t, x_{t-1}, .z_t, z_{t-1}...)$ and $(y_t, y_{t-1}, ...)$. For linear functions

$$\begin{aligned} & MSE[E(x_{t+s}|x_t, x_{t-1}, .z_t, z_{t-1}, .)] \\ = \quad & MSE[E(x_{t+s}|x_t, x_{t-1}, ..z_t, z_{t-1}.., y_t, y_{t-1}, .)] \end{aligned}$$

Granger Causality

Example

The Market efficiency hypothesis yields prices as a function of dividends:

$$P_t = \sum_{i=1}^{\infty} (1/(1+r))^i E(D_{t+i}|I_t)$$

Suppose

$$D_t = d + u_t + \delta u_{t-1} + \nu_t$$

where u_t and ν_t are independent white noise processes, then

$$E_t D_{t+i} = \begin{cases} d + \delta u_t & \text{for } i = 1 \\ d & \text{for } i = 2, 3, ... \end{cases}$$

Granger Causality

Example

The stock prices will be given by

$$P_t = d/r + \delta u_t/(1+r)$$

Thus,P_t is a white noise: no series should granger cause stock prices. Nevertheless, notice that:

$$\delta u_{t-1} = (1+r)P_{t-1} - (1+r)d/r$$

Substituting back in the D_t:

$$D_t = d + u_t + (1+r)P_{t-1} - (1+r)d/r + \nu_t$$

Thus stock prices Granger cause dividends

Granger Causality

Example

The bivariate VAR takes the form

$$\begin{bmatrix} P_t \\ D_t \end{bmatrix} = \begin{bmatrix} d/r \\ -d/r \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ (1+r) & 0 \end{bmatrix} \begin{bmatrix} P_{t-1} \\ D_{t-1} \end{bmatrix} + \begin{bmatrix} \delta u_t/(1+r) \\ u_t + v_t \end{bmatrix}$$

Hence in this model, Granger causation runs in the opposite direction from the true causation.

Impulse Response Functions

- Recall a stationary VAR has a VMA(∞) representation:

$$W_t = \sum_{z=0}^{\infty} \psi_z \varepsilon_{t-z} \quad , \psi_0 = I$$

- Lead the above expression s periods:

$$W_{t+s} = \sum_{z=0}^{\infty} \psi_z \varepsilon_{t+s-z}$$

- Evaluate the above expression at $z = s$. Then

$$\psi_s = \frac{\partial W_{t+s}}{\partial \varepsilon_t'}$$

has the interpretation of a dynamic multiplier

Impulse Response Functions

- $(\psi_s)_{ij}$ = effect of a one unit increase in the j^{th} variable's innovation at time t (ε_{jt}) for the value of the i^{th} variable at time $t+s$ ($W_{i,t+s}$), holding all other innovations at all dates constant
- You can find these multipliers numerically by simulation:
  - set $W_t = ... = W_{t-p} = 0$, then set $\varepsilon_{jt} = 1$ and all the other terms to zero, and simulate the system

$$W_t = A_1 W_{t-1} + A_2 W_{t-2} + ... + A_p W_{t-p} + \varepsilon_t$$

for $t, t+1, t+s$, with $\varepsilon_{t+1}, \varepsilon_{t+2}, ... = 0$ This simulation corresponds to the J column of the matrix ψ_s . By doing this for other values of j we get the whole matrix.

Impulse Response Functions

Definition

A plot of $(\psi_s)_{ij}$, that is row i column j of ψ_s , as a function of s is called the impulse response function. It describes the response of $W_{i,t+s}$ to a one time impulse in W_{jt} with all other variables dated t or earlier held constant.

Impulse Response Functions

- Define interim multipliers:

$$\sum_{j=1}^{m} \psi_j$$

- and the long run multiplier:

$$\sum_{j=1}^{\infty} \psi_j.$$

Impulse Response Function

- The assumption that a shock in one innovation does not affect others is problematic since

$$E(\varepsilon_t \varepsilon_t') = \Omega \neq \text{a diagonal matrix}$$

- Since Ω is symmetric and positive definite, it can be expressed as

$$\Omega = ADA'$$

where A is a lower triangular matrix and D is a diagonal Matrix.

Impulse Response Function

- Let $u_t = A^{-1}\varepsilon_t$, then

$$W_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \sum_{j=0}^{\infty} \psi_j A A^{-1} \varepsilon_{t-j} = \sum_{j=0}^{\infty} \psi_j^* u_{t-j}$$

where

$$\psi_j^* = \psi_j A$$

$$E(u_t u_t') = E(A^{-1}\varepsilon_t \varepsilon_t' (A^{-1})') = A^{-1}\Omega(A^{-1})' = A^{-1}ADA'(A^{-1})' = D$$

- The matrix D gives the variance of u_{jt}

Impulse Response Function

- A plot of ψ_s^* as a function of s is known as an orthogonalized impulse response function.
- The matrix

$$\psi_s^* = \frac{\partial W_{t+s}}{\partial u_t'}$$

  gives the consequences of an increase in W_{jt} by a unit impulse in u_t .
- Notice that

$$\psi_0^* = \psi_0 A = IA$$

  is lower triangular. This implies that the ordering of variables is of importance.
- The ordering cannot be determined with statistical methods.

Variance Decomposition

- Consider the error in forecasting a VAR s periods ahead:

$$W_{t+S} - \widehat{W}_{t+S|t} = \sum_{j=0}^{s-1} \psi_j \varepsilon_{t+s-j}, \qquad \psi_0 = I$$

- The mean squared error of this s-period ahead forecast is thus

$$MSE(\widehat{W}_{t+s|t}) = \Omega + \psi_1 \Omega \psi_1' + ... + \psi_{s-1} \Omega \psi_{s-1}'$$

- Let us now consider how each of the orthogonalized disturbances $(u_{1t}, ..u_{nt})$ contributes to this MSE.

Variance Decomposition

- Lets write

$$\varepsilon_t = Au_t = a_1 u_{1t} + ...a_n u_{nt},$$

  where a_j denotes the j^{th} column of the matrix A.
- Recalling that the u's are uncorrelated, we get

$$\Omega = a_1 a'_{1(n\times n)} Var(u_{1t}) + ... + a_n a'_{n(n\times n)} Var(u_{nt})$$

Variance Decomposition

- Substituting this in the MSE of the s period ahead forecast we get

$$MSE(\widehat{W}_{t+s|t}) = \sum_{j=1}^{n} Var(u_{jt})(a_j a_j' + \psi_1 a_j a_j' \psi_1' + ... + \psi_{s-1} a_j a_j' \psi_{s-1}')$$

- With this expression we can calculate the contribution of the j^{th} orthogonalized innovation to the MSE of the s-period ahead forecast.

$$Var(u_{jt})(a_j a_j' + \psi_1 a_j a_j' \psi_1' + \psi_2 a_j a_j' \psi_2' ... + \psi_{s-1} a_j a_j' \psi_{s-1}')$$

- Magnitude in general depends on the ordering of the variables

Structural VARs

- Blanchard (1989) considers the following structure

$$\varepsilon_{1t} = eu_{2t} + u_{1t}$$

$$\varepsilon_{2t} = c_{21}\varepsilon_{1t} + u_{2t}$$

where u_{1t} and u_{2t} are regarded as demand and supply shocks, while ε_{1t} and ε_{2t} are output and unemployment innovations respectively

Structural VARs

- Blanchard and Quah (1989) :
  - They argue that a demand shock should have a zero long-run effect while a supply shock will not.

$$\varepsilon_{1t} = a_1 u_{2t} + u_{1t}$$

$$\varepsilon_{2t} = a_2 u_{1t} + u_{2t}$$

    where the covariance of u_{jt} is assumed to be zero.

- Consider

$$W_t = A_1 W_{t-1} + A_2 W_{t-2} + ... + A_p W_{t-p} + \varepsilon_t$$

  is estimated and the implied MA representation is

$$W_t = \sum_{z=0}^{\infty} \psi_z \varepsilon_{t-z} \qquad , \psi_0 = I.$$

Structural VARs

- In terms of the shocks of interest, we will write

$$\varepsilon_t = A u_t$$

where A is now defined as

$$A = \begin{bmatrix} 1 & a_1 \\ a_2 & 1 \end{bmatrix}.$$

- The MA representation in terms of the u_t shocks becomes:

$$W_t = \sum_{z=0}^{\infty} \psi_z A u_{t-z} \qquad \text{, where } \psi_0 = I$$

- If the long run effect of a demand shock upon output say, W_{1t} , is to be zero:

$$[\sum_{z=0}^{\infty} \psi_z A]_{[1,1]} = 0$$

Structural VARs

- To see how we compute this restriction we first notice that

$$\sum_{z=0}^{\infty} \psi_z$$

  is a 2x2 Matix. Let the first row of this matrix be

$$[\delta_1, \delta_2]$$

- Then, the restriction is just

$$\delta_1 + a_2\delta_2 = 0 \text{ or } a_2 = -\delta_1/\delta_2.$$

- Thus one parameter can be found from this restriction

- The other three come from the fact that:

$$V(\varepsilon_t) = A \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} A'$$

  since there are three unknowns in $V(\varepsilon_t)$ to determine a_1, σ_1^2 and σ_2^2.
- All that is needed is to estimate δ_1, δ_2 .

- Notice that the long run multiplier is easy to compute, since we know that

$$\sum_{z=0}^{\infty} \psi_z = \psi(1)$$

$$\sum_{z=0}^{\infty} \psi_z L^i = \psi(L)$$

$$\sum_{z=0}^{\infty} \psi_z L^i = (I - A_1 L - A_2 L^2 - ... + A_p L^p)^{-1} = A(L)^{-1}$$

$$\psi(1) = (I - A_1 - A_2 - ..... + A_p)^{-1} = A(1)^{-1}$$

- All the information that is needed for the impulse response function, is obtained from the estimated parameters in the VAR.

Impulse Resonponse Functions using Local Projections

- It has been proposed in the literature an alternative way of carrying out impulse response functions by doing local projections. Assume you know you have a VAR(1) (which can be the companion form of a VAR(p)) of the type.

$$Y_t = AY_{t-1} + \varepsilon_t$$

Now instead of analysing the $MA(\infty)$ we substitute backwards τ times to get

$$Y_{t+\tau} = A\varepsilon_{t+\tau-1} + A^2\varepsilon_{t+\tau-2} + ... + A^\tau\varepsilon_t + A^{\tau+1}Y_{t-1} + \varepsilon_{t+\tau}$$

or

$$Y_{t+\tau} = A^{\tau+1}Y_{t-1} + u_{t+\tau}$$

where

$$u_{t+\tau} = A\varepsilon_{t+\tau-1} + A^2\varepsilon_{t+\tau-2} + ... + A^\tau\varepsilon_t + \varepsilon_{t+\tau}$$

- Then the coefficient of a regression of regression of $Y_{t+\tau}$ on Y_{t-1} has the interpretation of $\frac{\partial Y_{t+\tau}}{\partial \varepsilon_{t-1}}$, that is the the $\tau+1$ impulse response

- Points to consider

1) The shock needs to be identify. We can simply do an orthogonalization. $\frac{\partial Y_{t+\tau}}{\partial B\varepsilon_{t-1}} = A^{\tau+1}B^{-1}$.

where $B\varepsilon_{t-1}$ is the orthogonal shock.
2) If It is a VAR(1), then estimating

$$Y_{t+\tau} = A^{\tau+1}Y_{t-1} + u_{t+\tau}$$

give unbiased but probably very imprecise estimates of $A^{\tau+1}$, with standard errors that need to be corrected (all that can be done using the theoretical structure of $u_{t+\tau}$).

Non-Stationary Stochastic Time Series Models

Series de Tiempo

UTDT

January 2020

Non-Stationary Series

Importance of cheking for the existence of Unit Roots:

- Standard assymptotical theory is not valid
- Possibility of Spurrious Regressions.
- Regress series of the same order of integration.

STRUCTURE OF THE LECTURE

- Definition and Properties.
- Estimation
- Testing for Unit Roots.

Time series

- There are three important types of time series which one is likely to find in financial econometrics:

Stationary, $I(0)$
Trend stationary
Non-stationary, $I(1)$

- We have focused until now on stationary processes

Trend stationary

Definition

A trend stationary variable is a variable whose mean grows around a fixed trend. This provides a classical way of describing an economic time series which grows at a constant rate. A trend-stationary series tends to evolve around a steady, upward sloping curve without big swings away from that curve. For simplicity assume that the following process.

$$y_t = \alpha + \mu t + \varepsilon_t \text{ where } \varepsilon_t \tilde{} N(0, \sigma^2)$$

Notice that the mean of this process varies with time but the variance is constant.

$$E(y_t) = \alpha + \mu t$$

$$V(y_t) = E(\alpha + \mu t + \varepsilon_t - (\alpha + \mu t))^2 = \sigma^2$$

Remark: $y_t^* = y_t - (\alpha + \mu t)$ then y_t^* is stationary.

A Non stationary Series: I(1) Processes

Definition

An autoregressive process of order p, AR(p), has a **unit root** *if the polynomial in L, $(1 - \phi_1 L - ...\phi_p L^p)$ has a root equal to one. The simplest example of a process with a unit root is a random walk, i.e.,*

$$y_t = y_{t-1} + \varepsilon_t$$

where ε_t is i.i.d. with zero mean and constant variance.

- We can easily see that the variance of this processes does not exist: lagging the process one period we can write

$$y_{t-1} = y_{t-2} + \varepsilon_{t-1},$$

and substituting back we get

$$y_t = y_{t-2} + \varepsilon_{t-1} + \varepsilon_t.$$

Then, repeating this procedure we can easily show that

$$y_t = y_0 + \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t$$

- The mean can be calculated assuming that y_0 is fixed, then the mean is constant over time

$$E(y_t) = E(y_0 + \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t) = y_0$$

- The variance of y_t, "conditional" on knowing y_0, can be computed as

$$\begin{aligned} V(y_t) &= V(y_0 + \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t) \\ &= V(\varepsilon_1) + V(\varepsilon_2) + ... + V(\varepsilon_{t-1}) + V(\varepsilon_t) = t\sigma^2 \end{aligned}$$

  As we move further into the future this expression becomes infinite. We conclude that the variance of a unit root process is infinite.

Comments

- A unit root process will only cross the mean of the sample very infrequently
- A process that has a unit root is also called **integrated of order one**, denoted as $I(1)$
- A stationary process is an **integrated of order zero** process, denoted as $I(0)$
- *Why is this important?*
  - The study of econometric models with non-stationary data has been one of the most important concerns of econometricians in the last 30 years. The topic is very vast and we just will *mention* some of the most important issues.

Spurious Regressions

- Granger and Newbold (1974) have shown that, using $I(1)$, you can obtain an apparently significant regression (say with a high R^2) even if the regressor and the dependent variable are independent.
- They did this by generating independent random walk series against each other
- *Rule of thumb*: whenever you obtain a very high R^2 and a very low DW you should suspect that the result are spurious.
    - the reason for the low DW will be understood in what is to come

Regressing Series that are integrated of the same order

Caution!

You should always check that the variables you are regressing one against the other are of the same order if you are to obtain meaningful results

Testing for unit roots

- We will show that the standard t- test cannot be applied to a process with a unit root
  - under the null of $\alpha = 1$ we get a degenerated distribution
- We will find a distribution for $\widehat{\alpha}$ under this circumstances and see that:
  1. it is not t - student
  2. it is biased to the left

Testing for Unit Roots (II)

Consider the following model:

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

where ε_t is assumed to be N(0,σ^2). It can easily be shown that asymptotically

$$\sqrt{T}(\hat{\alpha}_T - \alpha) \quad \underrightarrow{L} \quad N(0, (1 - \alpha^2))$$

If we want to use this distribution for testing the null hypothesis that $\alpha = 1$, then we find that the distribution under the null "degenerates" (collapses in one point).

- To obtain a non-degenerate asymptotic distribution for $\hat{\alpha}_T$ in the unit root case, it turns out that we have to multiply by T and not by the square root of T. Then the unit root coefficient converges at a faster rate T than for the stationary case.
- To get a better sense of why scaling by T is necessary when the true value of α is unity consider the OLS estimate:

$$\hat{\alpha}_T = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2}.$$

Then, substituting y_t by the AR(1) process we get that:

$$\hat{\alpha}_T - \alpha = \frac{\sum_{t=1}^{T} y_{t-1} \varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

- Multiplying in both sides by T, we get

$$T(\hat{\alpha}_T - \alpha) = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2}.$$

- Now, under the null that $\alpha = 1$, y_t can be written as:

$$\begin{aligned} y_t &= y_0 + \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t \\ &= \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t \qquad \textit{if we assume } y_0 = 0. \end{aligned}$$

- Then under the null that $\alpha = 1$, y_t ~N$(0, \sigma^2 t)$.

- Now we proceed to find the distribution of the numerator: under the null,

$$y_t^2 = (y_{t-1} + \varepsilon_t)^2 = y_{t-1}^2 + 2y_{t-1}\varepsilon_t + \varepsilon_t^2$$

and rearranging terms we obtain

$$y_{t-1}\varepsilon_t = \frac{1}{2}(y_t^2 - y_{t-1}^2 - \varepsilon_t^2).$$

- Then, the sum which appears in the *numerator* can be expressed as

$$\sum_{t=1}^{T} y_{t-1}\varepsilon_t = \sum_{t=1}^{T} \frac{1}{2}(y_t^2 - y_{t-1}^2 - \varepsilon_t^2) = \frac{1}{2}(y_T^2 - y_0^2) - \sum_{t=1}^{T} \frac{1}{2}\varepsilon_t^2.$$

- Then, recalling that $y_0 = 0$ and multiplying by (T^{-1}) we obtain the expression of the *numerator* as the sum of two terms

$$T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t = (\frac{1}{2T})y_T^2 - \sum_{t=1}^{T}(\frac{1}{2T})\varepsilon_t^2$$

- To find the distribution of this expression we divide each side by σ^2 which yields the following result

$$(\sigma^2 T)^{-1} \sum_{t=1}^{T} y_{t-1}\varepsilon_t = (1/2)(\frac{y_T}{\sigma\sqrt{T}})^2 - \sum_{t=1}^{T}(\frac{1}{2\sigma^2 T})\varepsilon_t^2$$

- Consider the first term of this expression. Since we have shown above that y_t ~N$(0, \sigma^2 t)$, standardizing we obtain

$$(y_T/\sigma\sqrt{T})\tilde{}N(0,1),$$

- and then squaring this expression we find that the first term of the numerator is distributed Chi-square

$$(y_T/\sigma\sqrt{T})^2\tilde{}\chi^2(1).$$

- It can be shown using the law of large numbers that the second term converges in probability to σ^2, i.e.

$$(1/T)\sum_{t=1}^{T}\varepsilon_t^2 \quad \underrightarrow{P} \quad \sigma^2, \qquad \textit{or} \qquad (1/\sigma^2 T)\sum_{t=1}^{T}\varepsilon_t^2 \quad \underrightarrow{P} \quad 1$$

- If we put both results together we can see that the numerator converges to

$$(\sigma^2 T)^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \qquad \underrightarrow{L} \qquad (1/2)(X-1) \quad \textit{where} \quad X \tilde{} \chi^2(1).$$

- It can also be shown using the law of large numbers that the *denominator* converges in probability to

$$E(T^{-2}\sum_{t=1}^{T} y_{t-1}^2).$$

- Now,as y_{t-1} ~N$(0, \sigma^2(t-1))$, then $E(y_{t-1}^2) = \sigma^2(t-1)$. Therefore the expected value of the denominator can be written as

$$\begin{aligned} E(T^{-2}\sum_{t=1}^{T} y_{t-1}^2) &= T^{-2}\sum_{t=1}^{T} E(y_{t-1}^2) \\ &= T^{-2}\sigma^2 \sum_{t=1}^{T}(t-1) \\ &= \sigma^2 T^{-2}(T-1)T/2 \end{aligned}$$

- Then if we multiply $(\hat{\alpha}_T - 1)$ by T instead than by $\sqrt{T}$, we obtain a non-degenerate asymptotic distribution, but this distribution is not Gaussian.

How do we test for unit roots? Dicky Fuller Test

Consider the following model

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

where ε_t "is assumed" to be N(0,σ^2)

We want to test the Hypothesis of the existence of a unit root therefore we set the following null and alternative hypothesis.

$$\begin{aligned} H_0) \quad \alpha &= 1(UR) \\ H_1) \quad \alpha &< 1(I(0)) \end{aligned}$$

The obvious estimator of is the OLS estimator, $\hat{\beta}$. The problem is that under the null hypothesis there is considerable evidence of the non - adequacy of the asymptotic (approximate in large samples) distribution. Therefore
The latter equation can be rewritten as

$$\Delta y_t = \mu + (\alpha - 1) y_{t-1} + \beta t + \varepsilon_t$$

or

$$\Delta y_t = \mu + \lambda y_{t-1} + \beta t + \varepsilon_t$$

For this expression the relevant hypothesis should be written as:

$$\begin{aligned} H_0)\ \lambda &= 0\ (\textit{unit root}) \\ H_1)\ \lambda &< 0\ (I(0)) \end{aligned}$$

Fuller (1976) tabulated, using Monte Carlo methods, critical values for alternative cases, for example for a sample size of 100 the 5 % critical values are

$\mu = 0,\ \beta = 0$	-2.24
$\mu \neq 0,\ \beta = 0$	-3.17
$\mu \neq 0,\ \beta \neq 0$	-3.73

Therefore the method simply consist to check the t-statistic of $\hat{\lambda}$ against the critical values of Fuller (1976). Notice that the critical values depend on

i) the sample size

ii) whether you include a constant and/or a time trend.

This procedure is only valid when there is no evidence of serial correlation in the residuals, $\hat{\varepsilon}_t$ .If there is serial correlation you should need to include additional lags, say $\Delta y_{t-1}, \Delta y_{t-2}$,... until it disappears

Augmented Dickey Fuller

Consider:

$$\Delta y_t = \mu + \lambda y_{t-1} + \beta t + \alpha_1 \Delta y_{t-1} + ... + \alpha_k \Delta y_{t-k} + \varepsilon_t$$

- We chose to augment the regression with k lags. This is usually denoted as ADF(k).
- Choosing the order of augmentation of the DF regression:
  - *choosing k as a function of the number of observations as in Schwert (1989)*

$$k = INT(12(T/100)^{1/12})$$

  - *information based rules such as AIC and BIC.*
  - *sequential rules*
  - General to specific seems to be preferable to the other methods.

Small sample properties of the Dickey Fuller test

- The power of the test is extremely low
- ADF comes up with conflicting results depending on the order chosen for the regression
- Good practice: start with quite a general model and then delete lags

Phillips - Perron type tests for unit roots

- Alternative approach to DF
- They make a non-parametric correction to the standard deviation which provides a consistent estimator of the variance.

$$S_{Tl}^2 = T^{-1} \sum_{t=1}^{T} (\varepsilon_t^2) + 2T^{-1} \sum_{t=1}^{l} \sum_{t=j+1}^{T} \varepsilon_t \varepsilon_{t-j}$$

$$S_{\varepsilon}^2 = T^{-1} \sum_{t=1}^{T} (\varepsilon_t^2)$$

where $_l$ is the lag truncation parameter used to ensure that the autocorrelation of the residuals is fully captured.

An asymptotically valid test $\phi = 1$, for

$$\Delta y_t = \mu + (\phi - 1) y_{t-1} + \varepsilon_t \qquad \varepsilon_t \tilde{} iiD(0, \sigma^2)$$

when the underlying DGP is not necessarily an AR(1) process, is given by the Phillips Z-test.

$$Z(\tau_\mu) = (S_\varepsilon / S_{Tl}) \tau_\mu - (1/2)(S_{Tl}^2 - S_\varepsilon^2)[S_{Tl}[T^2 \sum_{t=2}^{T} (y_t - \overline{y})^2]^{.5}]^{-1}$$

where τ_μ is the t-statistic associated with testing the null hypothesis $\rho = 1$.

Comments The critical values for this test statistic are the same as those used for the same case in the fuller table.
Monte Carlo work suggests that the Phillips-type test has poor size properties (tendency to over reject when is true) when the underlying DGP has large negative MA components.

Structural breaks and unit roots

- Perron (1988) has shown that an I(0) process with a structural break in the mean will be difficult to distinguish from a I(1) process.
- If we know where the break takes place, the natural thing to do, is to partial out the break by using dummy variables and test for unit roots once the break has been partialled out.
- A possible solution to try to identify these breaks is to perform the *ADF* test recursively and to compute recursive t-statistics.

Recursive t-statistics

- The recursive ADF -statistic is computed using sub samples $t = 1..k$ *for* $k = k_0, ...T$, where k is the start up value and T is the sample size of the full sample.
- The most general model (with drift and trend) is estimated for each sub sample
- The minimum value of $\tau_\tau(k/T)$ across all the sub samples is chosen and compared with the table provided by Banjeree, Lumsdaine and Stock

Rolling ADF

- The previous method could also be applied using a (large enough) window (of size k) to see if there are clear changes in the pattern of a series.
- The most general model (with drift and trend) is estimated for each sub sample
- The minimum value of $\tau_\tau(k/T)$ across all the sub samples is chosen and compared with the table found in the class notes.

Tests with stationarity as a null: KPSS test

Consider the following model.

$$y_t = \alpha + \delta t + \xi_t + \varepsilon_t$$

where ε_t is a stationary process and ξ_t is a random walk given by

$$\xi_t = \xi_{t-1} + u_t \quad u_t \tilde{} iid(0, \sigma_u^2)$$

The null of stationarity is formulated as

$$H_0) \sigma_u^2 = 0$$

The test statistic for this hypothesis is given by

$$LM = \frac{\sum_{t=1}^{T} S_t^2}{\hat{\sigma}_e^2}$$

where e_t are the residuals of a regression of y_t on a constant and a time trend, $\hat{\sigma}_e^2$ is the residual variance for this regression and S_t is the partial sum of e_t defined by

$$S_t = \sum_{i=t}^{T} e_i \quad t = 1, 2, .., T.$$

Variance Ratio Tests

- This will provide us with another tool to discriminate between (trend) stationary and non-stationary series.
- Consider y_t and assume that it follows follows a random walk, i.e.

$$y_t = y_{t-1} + \varepsilon_t$$

then by iterative substitution we know that

$$y_t = y_{t-k} + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3} + ... + \varepsilon_{t-k+1}$$

- Denoting the difference between y_t and y_{t-k} as $\Delta_k y_t$, then

$$\Delta_k y_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3} + ... + \varepsilon_{t-k+1}$$

  and clearly the variance of $\Delta_k y_t$, is $\sigma^2 k$, where σ^2 is the variance of ε.
- We can define a "variance ratio" function (a function of k) as

$$\lambda_1(k) = \frac{Var(\Delta_k y_t)}{Var(\Delta_1 y_t)} = k.$$

- Therefore a plot of λ_1 against k should be an increasing straight line

- Alternatively we may define a new function $\lambda_2(k)$ as

$$\lambda_2(k) = \lambda_1(k)/k$$

- If there is a unit root, $\lambda_2(k)$ tends to one when k tends to infinite.
- However if y_t *does not* contain a unit root it can be shown that the lim $\lambda_2(k)$ when k tends to infinite is equal to zero.

Proof.

(idea)Assume the following *AR(1)* process

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \qquad t = 1, ..., T$$

□

we have seen that by iterative substitution we can express this process as

$$y_t = \phi_1^k y_{t-k} + \phi_1^{k-1} \varepsilon_{t-(k-1)} + \phi_1^{k-2} \varepsilon_{t-(k-2)} + ... + \phi_1 \varepsilon_{t-1} + \varepsilon_t$$

Then subtracting y_{t-k} in both sides of the equation we get the following expression

$$y_t - y_{t-k} = (\phi_1^k - 1) y_{t-k} + \phi_1^{k-1} \varepsilon_{t-(k-1)} + \phi_1^{k-2} \varepsilon_{t-(k-2)} + ... + \phi_1 \varepsilon_{t-1} + \varepsilon_t$$

Proof.

(continues)the variance of $y_t - y_{t-k}$, $Var(\Delta_k y_t)$ is equal to □

$$V(y_t - y_{t-k}) = (\phi_1^k - 1)^2 V(y_{t-k}) + V(\sum_{j=0}^{k-1} \phi_1^j \varepsilon_{t-j})$$

Notice that

$$V(y_{t-k}) = V(y_t) = (1/(1-\phi_1^2))\sigma^2$$

and

$$V(\sum_{j=0}^{k-1} \phi_1^j \varepsilon_{t-j}) = ((1-\phi_1^{2k})/(1-\phi_1^2))\sigma^2,$$

Proof.

We can write the variance of Δy_{t-k} as

$$V(y_t - y_{t-k}) = (\phi_1^k - 1)^2(1/(1-\phi_1^2))\sigma^2 + ((1-\phi_1^{2k})/(1-\phi_1^2))\sigma^2$$

In the same way we can express for a stationary process the variance of the first difference of y_t, $(y_t - y_{t-1}) = (\phi_1 - 1)y_{t-1} + \varepsilon_t$.

$$V(y_t - y_{t-1}) = (\phi_1 - 1)^2 V(y_{t-1}) + \sigma^2 = (\phi_1 - 1)(1/(1-\phi_1^2))\sigma^2 + \sigma^2$$

then the variance ratio can be written as

$$\lambda_1(k) = \frac{(\phi_1^k - 1)^2(1/(1-\phi_1^2))\sigma^2 + (\ (1-\phi_1^{2k})/(1-\phi_1^2)\ )\sigma^2}{(\phi_1 - 1)^2(1/(1-\phi_1^2))\sigma^2 + \sigma^2}$$

□

Proof.

Then the limit of $\lambda_1(k)$ when k tends to infinite for a stationary process is

$$lim_{k\to\infty}\lambda_1(k) = \frac{1}{1-\phi_1}$$

which is a constant provided that $\phi_1 \neq 1$.
It should be clear from the previous result that the limit of $\lambda_2(k)$ equals 0 when k tends to infinite.

Trend stationary and difference stationary processes

A trend stationary variable may be written as

$$y_t = \alpha + \mu t + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + ...$$

Then y_{t-k} is simply

$$y_{t-k} = \alpha + \mu(t-k) + \varepsilon_{t-k} + \theta_1 \varepsilon_{t-1-k} + \theta_2 \varepsilon_{t-2-k} + \theta_3 \varepsilon_{t-3-k} + ...$$

The k^{th} difference can be obtained simply by subtracting the two above equations

$$\begin{aligned} y_t - y_{t-k} &= \mu k + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_{k-1} \varepsilon_{t-(k-1)} + (\theta_k - 1)\varepsilon_{t-k} + \\ &\quad (\theta_{k+1} - \theta_1)\varepsilon_{t-(K-1)} + ... + (\theta_{k+q} - \theta_q)\varepsilon_{t-q} + ... \\ &= \mu k + \sum_{j=0}^{k-1} \theta_j \varepsilon_{t-j} + \sum_{j=0}^{\infty} (\theta_{k+j} - \theta_j)^2 \varepsilon_{t-j} \end{aligned}$$

Then the variance of $y_t - y_{t-k}$ may be written as

$$V(y_t - y_{t-k}) = \sigma^2 \sum_{j=0}^{k-1} \theta_j^2 + \sigma^2 \sum_{j=0}^{\infty} (\theta_{k+j} - \theta_j)^2$$

From the previous equation we can see that when k tends to infinity, the variance of $\Delta_k y_t$ is equal to

$$V(\Delta_k y_t) = 2\sigma^2 \sum_{j=0}^{k-1} \theta_j^2$$

Now the first difference of a trend stationary process, Δy_t is

$$\begin{aligned} y_t - y_{t-1} &= \mu + \varepsilon_t + (\theta_1 - 1)\varepsilon_{t-1} + ... + (\theta_{k+1} - \theta_k)\varepsilon_{t-(k+1)} \\ &\quad + ... + (\theta_{k+q} - \theta_{k+q-1})\varepsilon_{t-q} + ... \end{aligned}$$

Then the variance of the first difference can be written as

$$\begin{aligned} V(y_t - y_{t-1}) &= V(\varepsilon_t + (\theta_1 - 1)\varepsilon_{t-1} + ... + (\theta_{k+1} - \theta_k)\varepsilon_{t-(k+1)} \\ &\quad + ... + (\theta_{k+q} - \theta_{k+q-1})\varepsilon_{t-(k+q)} + ...) \\ &= \sigma^2(1 + \sum_{j=0}^{\infty}(\theta_{j+1} - \theta_j)^2) \end{aligned}$$

The variance ratio should be

$$\lambda_1(k) = \frac{Var(\Delta_k y_t)}{Var(\Delta_1 y_t)} = \frac{\sum_{j=0}^{k-1} \theta_j^2 + \sigma^2 \sum_{j=0}^{\infty} (\theta_{k+j} - \theta_j)^2}{(1 + \sum_{j=0}^{\infty} (\theta_{j+1} - \theta_j)^2)},$$

and the limit when k tends to infinity is

$$\lim_{k \to \to \infty} \lambda_1(k) = \lim_{k \to \to \infty} \frac{Var(\Delta_k y_t)}{Var(\Delta_1 y_t)} = \lim_{k \to \to \infty} \frac{2 \sum_{j=0}^{k-1} \theta_j^2}{(1 + \sum_{j=0}^{\infty} (\theta_{j+1} - \theta_j)^2)}.$$

- The expression is a constant and might be greater or smaller than one depending on the θ_j values.
- We can distinguish between the two models by simply noting that under the random walk assumption λ_1 increases with k and that under the trend stationary assumption λ_1 tends to a constant.
- Alternatively we can consider λ_2 $Var(\Delta_k y_t)/k$. and note both, that when the model is a random walk this expression tends to 1 (see proof above), and that this ratio should tend to zero when k tends to infinity since $Var(\Delta_k y_t)$ is constant for the trend stationary model.

- **Sampling distribution of λ(k) under the Random Walk Hypothesis.**

$$H_0)\ \alpha = 1 \ or \quad y_t = \mu + y_{t-1} + \varepsilon_t \qquad\qquad \varepsilon_t \tilde{} IIDN(0,)$$

It can be shown that asymptotically under the null, $\sqrt{Tk}(\hat{\lambda}_2(k) - 1) \overset{a}{\sim} N(0, 2(k-1))$. Then tests of the null Hypothesis can be carried out on the standardized statistics.

Cointegration

Series de Tiempo

UTDT

January 2020

Cointegration

Importance of checking for cointegration:

- If variables do not cointegrate there is not a meaningful relationship between those variables, then Change the Theory.
- If variables do cointegrate we need to write the model correctly.

STRUCTURE OF THE LECTURE

- Definition and Properties.
- Valid representations if the variables cointegrate
- Testing for Cointegration.

Cointegration

- A Cointegration relationship captures the long-run relationship between the variables of interest.
- The aim behind cointegration is the detection and analysis of long run relationships amongst economic time series variables.
- Cointegration analysis provides a way of retaining both short-run and long-run information.
- Cointegration is sometimes thought to be a pre-requisite for the validity of some economic theory.

Definition

Consider two $I(1)$ processes $\{Y_t\}$ and $\{X_t\}$.Their linear combination maybe $I(1)$ or $I(0)$. X and Y are said to cointegrate if there exists (a_1, a_2) such that $a_1 Y + a_2 X \tilde{} I(0)$.

Example

Consider the following example where cointegration of Prices and Dividends is a necessary condition for markets efficiency in the Fama sense. Let us assume that stock prices might be written as

$$P_t = \sum_{i=1}^{\infty} (1/(1+r))^i E(D_{t+i}|I_t) + \varepsilon_t$$

where we assume that ε_t is an I(0) process

Example

Let also assume D_t follows a random walk which is a special case of an I(1) variable.

$$D_t = D_{t-1} + \nu_t.$$

Then, we may express stock prices as

$$P_t = (1/r)D_t + \varepsilon_t$$

Example

Given that D_t are integrated of order one, stock prices also are integrated of order one. If the theory is valid, $(1, -(1/r))$ is going to be a cointegrating vector since

$$(1, -(1/r)) \begin{bmatrix} P_t \\ D_t \end{bmatrix} = Z_t = \varepsilon_t$$

Notice that we assumed that ε_t was I(0), therefore if the theory holds dividends and prices should be cointegrated.

- There are cases where there exist a theoretical long run cointegration relationship dictated by the model. In the previous example is not enough that Prices and Dividends cointegrate, the cointegrating vector has to be $(1, -(1/r))$ for the theory to hold.

Example

Consider the following model:

$$x_t + \beta y_t = u_t$$

$$x_t + \alpha y_t = e_t$$

$$u_t = u_{t-1} + \varepsilon_{1t}$$

$$e_t = \rho e_{t-1} + \varepsilon_{2t} \quad \text{with } |\rho| < 1$$

$(\varepsilon_{1t}, \varepsilon_{2t})'$ is distributed identically and independently as a bivariate normal with $E(\varepsilon_{1t}) = E(\varepsilon_{2t}) = 0$, $var(\varepsilon_{1t}) = \sigma_{11}$, $var(\varepsilon_{2t}) = \sigma_{22}$ $cov(\varepsilon_{1t}\varepsilon_{2t}) = \sigma_{12}$

Solving for x_t and y_t form the above system with $\alpha \neq \beta$ gives

$$\begin{aligned} x_t &= \alpha(\alpha-\beta)^{-1}u_t - \beta(\alpha-\beta)^{-1}e_t, \\ y_t &= -(\alpha-\beta)^{-1}u_t + (\alpha-\beta)^{-1}e_t. \end{aligned}$$

Then:

1. Both x_t and y_t are integrated of order one i.e., $x_t \tilde{} I(1)$, $y_t \tilde{} I(1)$, since u_t is integrated of order one.
2. $x_t + \alpha y_t$ is $I(0)$ because e_t is stationary.
3. The cointegration vector is $(1, \alpha)$ and $x + \alpha y$ is the equilibrium relationship.

Proposition *In the bivariate case if the equilibrium condition exists, is unique.*

Proof.

Suppose that there exist two distinct co-integrating parameters α and γ such that $x + \alpha y$ and $x + \gamma y$ are both $\tilde{I}(0)$. This implies that $(\alpha - \gamma)y_t$ is also $I(0)$ because a linear combination of two I(0) variable is also I(0). But we know that for $\alpha \neq \gamma$, $(\alpha - \gamma)y_t$ $\tilde{}$I(1) therefore we have a contradiction unless $\alpha = \gamma$. □

Representation Theorems

- Consider the model in the first example, with $|\rho| < 1$.
- Whenever x_t and y_t are cointegrated, we can show, for the simple two variable model, that the system of two equations has the following representations:
  - **Error-correction representation**
  - **Vector autoregressive in levels representation**
  - **Moving-average representation.**

Error correction mechanism

- There is a strong empirical motivation for these models since they perform very well.
- This representation has the advantage that it keeps long run and short run information.
- Allows to rewrite RE models which variables are I(1) in terms of I(0) variables.

ECM Representation

I repeat the above model for convenience:

$$x_t + \beta y_t = u_t$$

$$x_t + \alpha y_t = e_t$$

$$u_t = u_{t-1} + \varepsilon_{1t}$$

$$e_t = \rho e_{t-1} + \varepsilon_{2t} \quad \text{with } |\rho| < 1$$

Lagging the equations and substracting the lagged values:

$$\Delta x_t + \beta \Delta y_t = \Delta u_t$$

$$\Delta x_t + \alpha \Delta y_t = \Delta e_t$$

and rewriting the equations we have

$$\begin{bmatrix} 1 & \beta \\ 1 & \alpha \end{bmatrix} \begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} \Delta u_t \\ \Delta e_t \end{bmatrix}$$

- Inverting the matrix and noting that

$$\begin{aligned} \Delta u_t &= \varepsilon_{1t} \\ \Delta e_t &= -(1-\rho)e_{t-1} + \varepsilon_{2t} \end{aligned}$$

we have

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \alpha & -\beta \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ -(1-\rho)e_{t-1} + \varepsilon_{2t} \end{bmatrix}$$

- The ECM representation would be:

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \beta(1-\rho)(e_{t-1}) \\ -(1-\rho)(e_{t-1}) \end{bmatrix} + \begin{bmatrix} \zeta_{1t} \\ \zeta_{2t} \end{bmatrix}$$

where

$$\begin{bmatrix} \zeta_{1t} \\ \zeta_{2t} \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \alpha & -\beta \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

VAR representation

Now, notice that:

$$\begin{aligned} \Delta u_t &= \varepsilon_{1t} \\ \Delta e_t &= -(1-\rho)e_{t-1} + \varepsilon_{2t} \\ &= -(1-\rho)(x_{t-1} + \alpha y_{t-1}) + \varepsilon_{2t} \end{aligned}$$

Then using the ECM representation

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \beta(1-\rho)(e_{t-1}) \\ -(1-\rho)(e_{t-1}) \end{bmatrix} + \begin{bmatrix} \zeta_{1t} \\ \zeta_{2t} \end{bmatrix}$$

This can be rewritten to obtain:

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \beta(1-\rho)(x_{t-1} + \alpha y_{t-1}) \\ -(1-\rho)(x_{t-1} + \alpha y_{t-1}) \end{bmatrix} + \begin{bmatrix} \zeta_{1t} \\ \zeta_{2t} \end{bmatrix}$$

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \beta(1-\rho) & \beta(1-\rho)\alpha \\ -(1-\rho) & -(1-\rho)\alpha \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \zeta_{1t} \\ \zeta_{2t} \end{bmatrix}$$

Notice that this VAR representation **is not** a VAR in the first differences. The VAR can be written as

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \beta(1-\rho) + (\alpha - \beta) & \beta(1-\rho)\alpha \\ -(1-\rho) & -(1-\rho)\alpha + (\alpha - \beta) \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \zeta_{1t} \\ \zeta_{2t} \end{bmatrix}$$

Moving Average Representation

Using the fact that

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \alpha & -\beta \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \Delta u_t \\ \Delta e_t \end{bmatrix},$$

and noting that,

$$\begin{bmatrix} \Delta u_t \\ \Delta e_t \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ (1-L)(1-\rho L)^{-1}\varepsilon_{2t} \end{bmatrix},$$

The MA representation is given by:

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \alpha & -\beta \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ (1-L)(1-\rho L)^{-1}\varepsilon_{2t} \end{bmatrix}.$$

- The MA polynomial has a root equal to 1 and therefore is not invertible
- There isn't a VAR in differences representation.

Cointergration and Trends

- Two variables that are I(1) may cointegrate to an trend stationary representation.

Examples

Consider the following procesess for X_t and Y_t :

$$\begin{aligned} X_t &= \mu_x + X_{t-1} + \nu_t, \\ Y_t &= \mu_y + Y_{t-1} + \varepsilon_t. \end{aligned}$$

X and Y are said to cointegrate if there exists (a_1, a_2) such that $a_1 X + a_2 Y \tilde{} I(0)$. From the above example is eassy to see that the variables may cointegrate to a trend stationary process since:

$$\begin{aligned} X_t &= \mu_x t + X_0 + \nu_t + \nu_{t-1} + ... + \nu_1, \\ Y_t &= \mu_y t + Y_0 + \varepsilon_t + \varepsilon_{t-1} + ... + \varepsilon_1. \end{aligned}$$

then we know that $a_1(\nu_t + \nu_{t-1} + ...+ \nu_1) + a_2(\varepsilon_t + \varepsilon_{t-1} + ...+ \varepsilon_1) \tilde{} I(0)$, but $(a_1\mu_x + a_2\mu_y)t$ and/or $(a_1 X_0 + a_2 Y_0)$ might be different from zero.

Cointegration in the n variable case

Definition

Consider a n$\times$1 vector of stochastic variables $X_t = (X_{1t}, X_{2t}, .., X_{n_t})$. We say that the elements of the vector are cointegrated of order (d, b), which we denote $X_t \sim \text{CI}(d, b)$ if

1. Each of the components of X_t are I(d).
2. There exists (at least) a vector such that $Z_t = \alpha' X_t$ is I$(d - b)$ for $d \geq b > 0$.

Then, α is called the cointegrating vector

Comments

1. If $d = b = 0$, then $\alpha' X_t = 0$ defines a long-run equilibrium relationship.
2. Notice that α is not unique.
3. If $n > 2$, there may be $r \leq n - 1$ linearly independent $(n \times 1)$ vectors $(\alpha_1, ..., \alpha_r)$ such that $A' X_t \sim I(0)$
4. The vectors $(\alpha_1, ..., \alpha_r)$ are not unique

Granger Representation Theorem

Theorem

Consider an n - vector time series X_t which satisfies:

$$\Phi(L)X_t = c + u_t$$

where $\Phi(L) = I_n - \sum_{i=1}^{P} \Phi_i L^i$, and u_t is a white noise with positive definite covariance matrix. It is assumed that det$[\Phi(z)]$= 0 which implies $|z| \geq 1$. Suppose that there exist exactly r cointegrating relationships among the elements of X_t . Then:

1. *there exists an $(n \times r)$ matrix A, of rank $r < n$ such that $A'X_t$ ~I(0).*
2. *ΔX_t has an MA representation given by $\Delta X_t = \mu + \Psi(L)u_t$ with*

$$A'\Psi(1) = 0,$$

where

$$\Psi(L) = I_n + \sum_{i=1}^{\infty} \Psi_i L^i$$

Theorem

(continues)

[3.]$\Phi(1) = BA'$ *There exist a VAR representation in levels and the determinant of the polynomial in* Φ *has a unit root. There exists an ECM representation*

Error Correction Representation

Transform the original VAR

$$\Phi(L)X_t = c + u_t,$$

using the following relationship

$$I_n - \sum_{i=1}^{p} \Phi_i L^i = I_n - (\sum_{i=1}^{P} \Phi_i)L - (I - L)\sum_{i=1}^{p-1} \Gamma_i L^i$$

where

$$\Gamma_i = -\sum_{i=j}^{p-1} \Phi_{i+1} \qquad for\ j = 1, ..., p-1$$

Then

$$(I_n - \sum_{i=1}^{p} \Phi_i L^i) X_t = \left(I_n - \left(\sum_{i=1}^{p} \Phi_i L + (I - L) \sum_{i=1}^{p-1} \Gamma_i L^i \right) \right) X_t = c + u_t .$$

Rearranging terms

$$\begin{aligned} X_t &= c + (\sum_{i=1}^{p} \Phi_i) X_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + u_t \\ &= c + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + (I_n - \Phi(1)) X_{t-1} + u_t \end{aligned}$$

$$\Delta X_t = c + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} - \Phi(1) X_{t-1} + u_t$$

Thus:

(a) If rank $[\Phi(1)] = 0$ then $\Phi(1) = 0$ and $X_t \tilde{} I(1)$

(b) If rank $[\Phi(1)] = n$ then $det(\Phi(1)) \neq 0$ ($\Phi(L)$ does not have a unit root). This implies that $X_t \tilde{} I(0)$

(c) If rank $[\Phi(1)] = r$, $0 < r < n$ then $\Phi(1) = BA'$, where B is an nxr matrix .

Restrictions on the parameters of the ECM representation

$$\Delta X_t = c + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} - BZ_{t-1} + u_t, \; Z_t = A' X_t$$

- $A'X_t$ gives the "error" from the long run equilibrium relationship, and B gives the "correction" to X_t
- Taking expected values in both sides we get;

$$[I - \sum_{i=1}^{p-1} \Gamma_i L^i] E(\Delta X_t) = c - BE(Z_{t-1})$$

- If we assume that there is no autonomous growth component

$$[I - \sum_{i=1}^{p-1} \Gamma_i L^i] E(\Delta X_t) = c - BE(Z_{t-1}) = 0$$

or

$$c = BE(Z_{t-1})$$

- Thus,

$$\begin{aligned} \Delta X_t &= BE(Z_{t-1}) + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + -BZ_{t-1} + u_t \\ &= -B(Z_{t-1} - E(Z_{t-1})) + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + u_t \end{aligned}$$

- The intercept enters the system only via the error-correction term and there is no autonomous growth component.

Tests for Cointegration: Univariate

Two stages Aproaches:First stage

- Consider the following model:

$$y_t = \beta x_t + \varepsilon_t$$

- And suppose that the actual model is the following dynamic model

$$y_t = \gamma_o x_t + \gamma_1 x_{t-1} + \alpha y_{t-1} + \varepsilon_t$$

which can be re-written as

$$y_t = \lambda_0 x_t + \lambda_1 \Delta x_t + \lambda_2 \Delta y_t + \varepsilon_t$$

where

$$\lambda_0 = \frac{\gamma_0 + \gamma_1}{1-\alpha}, \lambda_1 = \frac{-\gamma_1}{1-\alpha}, \lambda_2 = \frac{-\alpha}{1-\alpha}.$$

- Then the Static model, long-run parameter $\beta \Leftrightarrow$ dynamic model without the short run terms. The omitted dynamic terms are captured by the residuals and since the variables in levels converge at speed T, and those in differences at $.\sqrt{T}$, that won't bias the estimates of the static model.

Two stage approaches Second stage

As a second stage, we can use alternative testing strategies

1. **Make an (ADF) test for unit roots for the residuals**: *The Engle - Granger Approach.*
2. **The null may be tested using the Sargan-Bhargava or CDW test**

The Engle - Granger Approach

- If you do not reject the Hypothesis that the residuals have a unit root, then Y and X are not cointegrated.
- Regress

$$\Delta\hat{\varepsilon}_t = \phi\hat{\varepsilon}_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta\hat{\varepsilon}_{t-i} + \mu + \delta t + \zeta_t$$

- Standard Dickey - Fuller critical values: over reject the null of no cointegration. See MacKinon

The cointegrating Durbin Watson

- This test is very simple and consist in comparing the DW statistic with tabulated values.
- Consider equations :

$$x_t + \alpha y_t = e_t$$

$$e_t = \rho e_{t-1} + \varepsilon_{2t} \quad \text{with } |\rho| < 1$$

- A regression of x_t on y_t will yield serially correlated residuals. We can use the DW statistic to get information about ρ since this statistic is approximately $2(1-\rho)$.

$$DW \cong 2(1-\rho)$$

Three stages: The Engle - Granger - Yoo approach

- They assume a unique cointegrating vector and weak exogeneity of the short run parameters
- Third step: provides a correction for the first stage estimate of β
- They correct the long run relationship by the small sample bias $\frac{\gamma}{1-\alpha}$

The Johansen approach

- Define X_t, a vector of n potentially endogenous variables.
- It is possible to specify the following DGP and the model X_t as an unrestricted VAR involving p lags of X_t :

$$\Phi(L)X_t = c + u_t$$

- Using the ECM

$$\Delta X_t = c + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} - \Phi(1)X_{t-1} + u_t$$

where $\Gamma_i = -\sum_{i=j}^{p-1} \Phi_{i+1}$ *for* $j = 1, ..., p-1$, and $\Phi(1) = I_n - \sum_{i=1}^{P} \Phi_i$.

Consequently testing for cointegration amounts to a consideration of the rank of $\Phi(1)$, that is, finding the number of r linearly independent columns in $\Phi(1)$.

1. If $\Phi(1)$ is full rank the variables in X_t have to be I(0)
2. If $\Phi(1)$ is has zero rank there is no cointegrating vector
3. If $\Phi(1)$ is reduced rank the number of cointegrating vectors is the RANK($\Phi(1)$)

Canonical correlations

Population Canonical Correlations

Let the $(n_1 \times 1)$ vector y_t and the $(n_2 \times 1)$ vector x_t denote stationary random variables. In general

$$\begin{bmatrix} E(y_t y_t') & E(y_t x_t') \\ E(x_t y_t') & E(x_t x_t') \end{bmatrix} = \begin{bmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix}$$

We can often gain some insight into the nature of these correlations by defining two new $(n \times 1)$ random vectors φ_t and ξ_t , where n is the smaller of n_1 and n_2 .

$$\begin{aligned} \varphi_t &= K' y_t \\ \xi_t &= A' x_t \end{aligned}$$

The matrices K' and A' are chosen such that

$$\begin{bmatrix} E(y_t y_t') & E(y_t x_t') \\ E(x_t y_t') & E(x_t x_t') \end{bmatrix} = \begin{bmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix}$$

$$E(\varphi_t \varphi_t') = K' \Sigma_{YY} K = I$$

$$E(\xi_t \xi_t') = A' \Sigma_{XX} A = I$$

and

$$E(\varphi_t \xi_t') = R = \begin{bmatrix} r_1 & & 0 \\ & r_2 & \\ 0 & & r_n \end{bmatrix}$$

where the elements of φ_t and ξ_t are ordered in such a way that $1 \geq r_1 \geq r_2 ..... \geq r_n \geq 0$

Definition

The population parameter r_i is known as the i^{th} population canonical correlation between y_t and x_t .

Lemma

The canonical correlations can be calculated by calculating the eigen values of

$$\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY},$$

$\lambda_1 > \lambda_2 > ...\lambda_n$ *, and the canonical correlations in turn are the square roots of these eigenvalues.*

The Johansen Method of Reduced Rank Regression

A way to assess the rank of If $\Phi(1)$ is to use the auxiliary regressions which allow to find an equivalent method to maximizing the likelihood of the ECM model. This equivalent expression allows to test how many cointegrating relationships do exist.

1. Rewrite the ECM Equation

$$\Delta X_t + BA' X_{t-1} \quad = c + \Gamma_1 \Delta X_{t-1} + ... + \Gamma_{p-1} \Delta X_{t-(p-1)} + u_t$$

2. It is possible to correct for the short run dynamics by regressing ΔX_t and X_{t-1} separately on the right hand side of the previous equation

$$\begin{aligned} \Delta X_t &= c_1 + P_1 \Delta X_{t-1} + ... + P_{p-1} \Delta X_{t-(p-1)} + R_{0t} \\ X_{t-1} &= c_2 + T_1 \Delta X_{t-1} + ... + T_{p-1} \Delta X_{t-(p-1)} + R_{kt} \end{aligned}$$

3. The latter can be used to form the residual (product moment) matrices

$$\widehat{S}_{ij} = T^{-1} \sum \widehat{R}_{it} \widehat{R}'_{jt} \qquad i, j = 0, k$$

4. The maximum likelihood estimate of A is obtained as eigenvectors corresponding to the largest eigenvalues of:

$$|\lambda \widehat{S}_{kk} - \widehat{S}_{0k} \widehat{S}_{00}^{-1} \widehat{S}_{0k}| = 0$$

5. This gives n eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > .....\hat{\lambda}_n$ and their corresponding eigen vectors, $\hat{\varphi} = (\hat{\varphi}_1 > \hat{\varphi}_2 > ... > \hat{\varphi}_n)$

Testing for reduced rank

- The use of the auxiliary regressions allows to write the likelihood as a function of the eigen values. The likelihood that corresponds to the ECM without any restriction is

$$L^* = \begin{array}{c} -(Tn/2)log(2\pi) - (Tn/2) - (T/2)log\left|\widehat{S}_{00}\right| \\ -(T/2)\sum_{i=1}^{n}\log(1-\hat{\lambda}_i) \end{array}.$$

- To find the number of cointegrating vectors is equivalent to find the number of linearly independent columns in $\Phi(1)$ or the number of $n-r$ columns with small associated eigen values.
- Thus to test the null hypothesis that there are at most r cointegration vectors amounts to test

$$H_0)\lambda_i = 0 \quad i = r+1, ..., n.$$

where only the first r eigenvalues are non-zero.

- Likelihood that corresponds to the ECM without any restriction:

$$L^* = \begin{array}{c} -(Tn/2)log(2\pi) - (Tn/2) - (T/2)log\left|\widehat{S}_{00}\right| \\ -(T/2)\sum_{i=1}^{n}\log(1-\hat{\lambda}_i) \end{array}.$$

- Likelihood that corresponds to the ECM under the null that there are only r cointegrating vectors is:

$$L^*(H_0) = \begin{array}{c} -(Tn/2)log(2\pi) - (Tn/2) - (T/2)log\left|\widehat{S}_{00}\right| \\ -(T/2)\sum_{i=1}^{r}\log(1-\hat{\lambda}_i) \end{array}.$$

Trace statistic

- Thus to test the null hypothesis that there are at most r cointegration vectors amounts to test

$$H_0)\lambda_i = 0 \quad i = r+1, ..., n.$$

where only the first r eigenvalues are non-zero. (This is tested against the alternative of n cointegrating vectors).

- A likelihood ratio test, *using a non standard distribution*, can be constructed, using what is known as the **Trace statistic.**

$$\lambda_{trace} = -T \sum_{i=r+1}^{n} \log(1 - \hat{\lambda}_i) \qquad r = 0, 1, ..., n-2, n-1$$

Max statistic

Another test of the significance of the largest λ_r is the so called maximal-eigenvalue or $\boldsymbol{\lambda - \max}$ **statistic** :

$$\lambda_{\max} = -T \log(1 - \hat{\lambda}_{r+1}) \qquad r = 0, 1, ..., n-2, n-1.$$

This tests the existence of r cointegrating vectors against the alternative that $r + 1$ exist and is derived in exactly same way

Testing restrictions on the cointegrating vector

- Many times we are interested to test restrictions on the cointegrating vector.
- Therefore the asymptotic theory is standard since the regressions only involve variables which are I(0)and the test would be distributed *chi-square*. The LR test will be:

$$LR = -T \sum_{i=1}^{r} \log(1 - \hat{\lambda}_i) + T \sum_{i=1}^{r} \log(1 - \tilde{\lambda}_i)$$

ARCH Models

Series de Tiempo

UTDT

2020

- Different ARCH models are proposed to capture the "Stylized Facts" of Asset returns.

Uses of Univariate ARCH models:

- Forecasting Volatility: Need to find a model to characterize the series which then is used to produce a forecast.
- Forecasting volatility might be useful when pricing Derivatives.

STRUCTURE OF THE LECTURE

- Definition and Properties.
- Estimation and testing
- Multivariate extensions

Empirical Regularities of Asset Returns

- Thick tails : returns tend to be leptokurtic
- Volatility clustering
- Leverage effects
- Non Trading Periods
- Forecastable Events

ARCH (1)

Consider the following model:

$$y_t = \mu + \varepsilon_t$$

$$\varepsilon_t = v_t(\omega + \alpha\varepsilon_{t-1}^2)^{1/2}, v_t \tilde{} IIN(0,1), \omega > 0, \alpha > 0$$

NOTICE that $(\omega + \alpha\varepsilon_{t-1}^2)^{1/2}$ is the conditional standard deviation, σ_t defined as $(E(\varepsilon_t^2|I_{t-1}))^{1/2}$.

ARCH (1)

- $E(\varepsilon_t|I_{t-1}) = E(v_t|I_{t-1})(\omega + \alpha\varepsilon_{t-1}^2)^{1/2} = 0$
  - Since $E(v_t|I_{t-1}) = E(v_t) = 0$
- $Var(\varepsilon_t|I_{t-1}) = E(\varepsilon_t^2|I_{t-1}) = E(v_t|I_{t-1})(\omega + \alpha\varepsilon_{t-1}^2) = (\omega + \alpha\varepsilon_{t-1}^2)$
  - Since $E(v_t|I_{t-1}) = E(v_t) = 1$
- Then we define the first order ARCH model as

$$\sigma_t^2 = (\omega + \alpha\varepsilon_{t-1}^2)$$

- Notice that whenever the process is stationary applying iterative expectations we find that :

$$\begin{aligned} V(\varepsilon_t) &= E(\varepsilon_t^2) = E(\sigma_t^2) = E(\omega + \alpha\varepsilon_{t-1}^2) \\ E(\varepsilon_t^2) &= \omega/(1-\alpha) \end{aligned}$$

- An alternative representation
  We can always write the following expression

$$\begin{aligned} \varepsilon_t^2 &= E(\varepsilon_t^2 | I_{t-1}) + \eta_t, \text{ where } \eta_t \text{ is an innovation.} \\ & \quad \textit{or} \\ \varepsilon_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \eta_t \end{aligned}$$

  then if $0 < \alpha < 1$, we find that

$$E(\varepsilon_t^2) = \omega / (1 - \alpha)$$

ARCH (1)

Consider Now the properties of $y_t = \mu + \varepsilon_t$ when there are ARCH(1) effects:

- $E(y_t|y_{t-1}) = \mu$
- $Var(y_t|y_{t-1}) = (\omega + \alpha\varepsilon_{t-1}^2)$
- $V(y_t) = V(\varepsilon_t) = \omega/(1-\alpha)$
  (since $V(\varepsilon_t) = E(\varepsilon_t^2) = E(\omega + \alpha\varepsilon_{t-1}^2) = \omega + \alpha E(\varepsilon_{t-1}^2)$)

First order autoregressive process with ARCH effects

- Consider:

$$y_t = \theta y_{t-1} + \varepsilon_t$$

where $\varepsilon_t = v_t(\omega + \alpha\varepsilon_{t-1}^2)^{\frac{1}{2}}$ and $v_t \sim IIN(0,1), \omega > 0, \alpha > 0$

- $E(\varepsilon_t|I_{t-1}) = E(v_t|I_{t-1})(\omega + \alpha\varepsilon_{t-1}^2)^{\frac{1}{2}} = 0 \ \ (E(v_t|I_{t-1}) = E(v_t) = 0)$

First order autoregressive process with ARCH effects

- Conditional variance:

$$Var(\varepsilon_t|I_{t-1}) = E(v_t|I_{t-1})(\omega + \alpha\varepsilon_{t-1}^2) = (\omega + \alpha\varepsilon_{t-1}^2)$$

- Conditional mean and variance of y_t :

$$E(y_t|y_{t-1}) = \theta y_{t-1}$$

$$Var(y_t|y_{t-1}) = (\omega + \alpha\varepsilon_{t-1}^2)$$

First order autoregressive process with ARCH effects

- Unconditional variance of y_t:

$$Var(y_t) = E(Var(y_t|y_{t-1})) + Var(E(y_t|y_{t-1}))$$

- Thus:

$$E(Var(y_t|y_{t-1})) = E(\omega + \alpha\varepsilon_{t-1}^2) = \omega + \alpha E(\varepsilon_{t-1}^2) = \omega + \alpha Var(\varepsilon_{t-1})$$

$$Var(E(y_t|y_{t-1})) = \theta^2 Var(y_{t-1})$$

then

$$\begin{aligned} Var(y_t) &= \omega + \alpha Var(\varepsilon_{t-1}) + \theta^2 Var(y_{t-1}) \\ &\quad \textit{if stationary} \\ &= \omega + \alpha \frac{\omega}{(1-\alpha)} + \theta^2 Var(y_{t-1}) \\ &= \frac{\omega}{(1-\alpha)} + \theta^2 Var(y_{t-1}) \\ &\quad \frac{\omega}{(1-\alpha)(1-\theta^2)} \end{aligned}$$

ARCH Models

Definition

An ARCH (q) model for the time varying conditional variance is

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 = w + \alpha(L)\varepsilon_{t-1}^2$$

Remark A sufficient condition for the conditional variance to be positive is that the parameters of the model satisfy: $\omega > 0, (\forall i)\alpha_i \geq 0$

ARCH Models

Defining $\nu_t \equiv \varepsilon_t^2 - \sigma_t^2$, the $ARCH(q)$ model can be re-written as

$$\varepsilon_t^2 = \omega + \alpha(L)\varepsilon_{t-1}^2 + \nu_t$$

(Notice that $\sigma_t^2 = E(\varepsilon_t^2 | I_{t-1})$)

- AR(q) model for the squared innovations, ε_t^2 .
- Covariance stationary $\Leftrightarrow$ the sum of the positive autoregressive parameters is less than one, which gives:

$$Var(\varepsilon_t^2) = \omega / (1 - \alpha_1 - \alpha_2 ... - \alpha_q).$$

- ε_t 's are clearly not independent through time.

GARCH Models

- ARCH(q): too many parameters,
  $\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \sum_{i=2}^{q} \alpha_i \varepsilon_{t-i}^2$
- Generalized ARCH or GARCH (1,1):

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

- Generalized ARCH or GARCH (p,q):

$$\begin{aligned} \sigma_t^2 &= \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \\ &= \omega + \alpha(L) \varepsilon_{t-1}^2 + \beta(L) \sigma_{t-1}^2 \end{aligned}$$

- A sufficient condition for the conditional variance in the model to be well defined is that all the coefficients in the infinite order linear ARCH are positive

GARCH Models

- Rearranging the GARCH(p,q) model by defining

$$\nu_t = \varepsilon_t - \sigma_t^2$$

it follows that

$$\varepsilon_t^2 = \omega + (\alpha(L) + \beta(L))\varepsilon_{t-1}^2 - \beta(L)\nu_{t-1} + \nu_t$$

- ARMA(Max(p, q),p) model for ε_t^2

GARCH Models

- Covariance stationary $\Leftrightarrow$ all the coefficients of $(1 - \alpha(L) - \beta(L))$ lie outside the unit circle
- $\Leftrightarrow$ sum of the autoregressive coefficients is less than one
- Standard time series techniques in the identification of the orders of p and q

Persistence and stationarity

- GARCH (1,1)
- Assuming $\alpha + \beta < 1$, the unconditional variance of ε_{t+1} is:

$$\frac{\omega}{1 - (\alpha + \beta)}$$

- We can rewrite the GARCH (1,1) as:

$$\varepsilon_t^2 = \omega + (\alpha + \beta)\varepsilon_{t-1}^2 - \beta\nu_{t-1} + \nu_t$$

Persistence and stationarity

- The conditional expectation of volatility j periods ahead is:

$$E_{t-1}[\sigma^2_{t+j}] = \frac{w}{(1-\alpha-\beta)} + (\sigma^2_t - \frac{w}{(1-\alpha-\beta)})(\alpha+\beta)^j$$

Since, using iterative expectations,
$E_{t-1}[\varepsilon^2_{t+j}] = E_{t-1}[E_{t+j-1}\varepsilon^2_{t+j}] = E_{t-1}[\sigma^2_{t+j}]$

IGARCH Models

Definition

Integrated GARCH models are processes were the autorregresive part of the square residuals has a unit root, i.e., $(\alpha + \beta) = 1$. For this case the conditional expectation of the volatility j periods ahead is

$$E_t[\sigma^2_{t+j}] = \sigma^2_t + j\omega.$$

- Looks very much like a random walk with drift ω.
- The unconditional variance does not exist

EGARCH Models

- GARCH models not well suited to capture the "leverage effect"
- In the exponential GARCH (EGARCH) σ_t^2 depends on both the size and the sign of lagged residuals.

EGARCH Models

- EGARCH(1,1):

$$\ln \sigma_t^2 = \alpha_0 + \beta_1 \ln \sigma_{t-1}^2 + \gamma_0([|\varepsilon_{t-1}/\sigma_{t-1}| - (2/\pi)^{1/2}] + \delta(\varepsilon_{t-1}/\sigma_{t-1}))$$

- Obviously the EGARCH model always produces a positive conditional variance σ_t^2 for any choice of α_0, β_1, γ_0
- Because of the use of both $|\varepsilon_t/\sigma_t|$ and (ε_t/σ_t), σ_t^2, it will also be non-symmetric in ε_t and, for negative δ, will exhibit higher volatility for large negative ε_t.

Other ARCH Specifications

Glosten, Jagannathan and Runkle (1989) proposed the following specification:

$$\varepsilon_t = \sigma_t \nu_t$$

where ν_t is iid.

$$\sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-1}^2 I_{t-1},$$

where, $I_{t-1} = 1$ if $\varepsilon_{t-1} \geq 0$ and $I_{t-1} = 0$ if $\varepsilon_{t-1} < 0$.

Additional Explanatory Variables:

It is straightforward to add other explanatory variables to a GARCH specification:

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2 + \gamma X_{t-1},$$

where X is any positive variable.

GARCH-in-Mean Models

- Many theories in finance assume some kind of relationship between the mean of a return and its variance .
- *GARCH in Mean Models* allow for the conditional variance to have mean effects.
- Time varying risk premium.

GARCH-in-Mean Models

- Consider the following model:

$$y_t = \theta x_t + \psi \sigma_t^2 + \varepsilon_t$$

and

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_{t-1}^2 + \beta(L)\sigma_{t-1}^2$$

ARCH-in-Mean Models

An ARCH in Mean model simply models the conditional variance as an ARCH model instead of modeling as GARCH, i.e.

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_{t-1}^2$$

ARCH-in-Mean Models

Example

ARCH(1)-M

Consider a simple version of the above model:

$$y_t = \mu + \lambda\sigma_t^2 + \varepsilon_t$$

where $\varepsilon_t = v_t\sigma_t \quad v_t$ ~N(0,1)

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2$$

ARCH-in-Mean Models

Example

Then y_t may be expressed as

$$y_t = \mu + \lambda(\omega + \alpha\varepsilon_{t-1}^2) + \varepsilon_t$$

Then the expected value of y_t is

$$E(y_t) = \mu + \lambda\omega + \lambda\alpha E(\varepsilon_{t-1}^2)$$

and using that $E(\varepsilon_{t-1}^2) = \omega/(1-\alpha)$then

$$E(y_t) = \mu + \lambda(\omega/(1-\alpha))$$

Lagrange Multiplier test

- Check there are ARCH effects in the residuals
- Lagrange Multiplier Test for ARCH :

  Under the null, the AR(p) process for y_t

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t$$

  where ε_t is a Gaussian white noise process, $\varepsilon_t | I_{t-1}$ ~N$(0,\sigma^2)$

Lagrange Multiplier Test

The test for ARCH(q) effect simply consists on regressing

$$\hat{\varepsilon}_t^2 = w + \alpha_1 \hat{\varepsilon}_{t-1}^2 + \alpha_2 \hat{\varepsilon}_{t-2}^2 + ... + \alpha_q \hat{\varepsilon}_{t-q}^2 + \psi_t$$

Under the null hypothesis that $\alpha_1 = \alpha_2 = ... = \alpha_q = 0$, TR^2 is asymptotically distributed $\chi(q)$, where T is the number of observations.

Lagrange Multiplier Test

Caution! While this is the most widely used test we should be careful in interpreting the results. If the model is misspecified it is quite likely to reject the null

- Breaks in the variance will look as ARCH effects when using the whole sample
- In these cases, it is better to divide the sample and test for ARCH effects in the subperiods

Problem: How to divide the sample ex ante

Garch Effects and Sampling Frequency

- GARCH models do not temporaly aggregate
- Difficult to determine at which sampling frequency the data presents GARCH effects
- Empirical regularity: the higher is the sampling frequency, the higher the GARCH effects

Maximum Likelihood Estimation with Gaussian Errors

The estimation of GARCH type models is easily done by conditional maximum likelihood.
If the model to be estimated is

$$y_t = x_t\theta + \varepsilon_t$$

where ε_t ~N(0,σ_t^2), and the conditional variance is assumed to be GARCH(1,1), i.e. ;

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

Then the conditional distribution of y_t is

$$f(y_t|x_t, I_{t-1}) = (2\pi\sigma_t^2)^{-.5}\exp(-.5(y_t - x_t\theta)^2/\sigma_t^2)$$

Maximum Likelihood Estimation with Gaussian Errors

Conditional log likelihood is

$$\log L(\theta, \omega, \alpha, \beta | I_{t-1}) = \sum_{t=1}^{T} (-.5 \log(2\pi) - .5 \log(\sigma_t^2) - .5\sigma_t^{-2}(y_t - x_t\theta)^2)$$

Notice that at time 1 we need initial values for ε_0^2 and σ_0^2:

$$\sigma_0^2 = \omega/(1 - \alpha - \beta)$$

and

$$\varepsilon_0^2 = (\omega/(1 - \alpha - \beta))$$

Maximum Likelihood Estimation with Non-Gaussian Errors

- Some unconditional distributions seem to have fatter tails than the normal:
  - another distribution for ε_t : t-distribution.

$$f(\varepsilon_t) = \frac{(\Gamma[(\nu+1)/2]/\Gamma(\nu/2))((\nu-2)\pi\sigma_t^2)^{-.5}}{[1+(\varepsilon_t^2/(\sigma_t^2(\nu-2))]^{-(\nu+1)/2}}$$

n represents the degrees of freedom (estimate)

How to compare between Different GARCH Specifications

- Most GARCH models are non nested
  - this makes comparison not straight forward

Misspecification tests on the standardized residuals

- ARCH(1):

$$\varepsilon_t = v_t(\omega + \alpha\varepsilon_{t-1}^2)^{1/2}$$

- Test for the existence ARCH effects in the standardized residuals

$$\hat{v}_t = \hat{\varepsilon}_t / (\hat{\omega} + \hat{\alpha}\hat{\varepsilon}_{t-1}^2)^{1/2}$$

Comparison between alternative models based on auxiliary regressions

- Auxiliary regression as a mean of choosing between different ARCH models:

$$\hat{\varepsilon}_t^2 = \alpha + \beta \hat{\sigma}_t^2 + \xi_t$$

- Regress the squared residuals on the fitted variance of the alternative GARCH models.
- Pagan and Scwhert (1989) propose to test the joint hypothesis:

$$H_0) \ \alpha = 0, \beta = 1 \qquad H_1) \ \alpha \neq 0, \beta \neq 1$$

Comparison between alternative models based on auxiliary regressions

- As a second step, they propose to compare the models that were not rejected on the basis of goodness of fit.
- The better the fit the better it mimics the conditional variance.
- Scale effects: logs

Measuring the accuracy of Forecasts of Different ARCH Models

- Forecasting ability of the different ARCH models as a way of comparing them
- Various *loss functions* have been proposed :
  - *Mean Squared Error*

$$MSE = (1/l)(\sum_{t=T}^{T+l}(\hat{\varepsilon}_t^2 - \hat{\sigma}_t^2)^2)$$

  - *Mean Absolute Error.*

$$MAE = (1/l)(\sum_{t=T}^{T+l}|\hat{\varepsilon}_t^2 - \hat{\sigma}_t^2|)$$

- *Mean Squared Error of the log of the squared residuals.*

$$[LE]^2 = (1/I)(\sum_{t=T}^{T+I}(\ln(\hat{\varepsilon}_t^2) - \ln(\hat{\sigma}_t^2))^2)$$

- *Mean Absolute Error of the log of the squared residuals.*

$$[MAE]^2 = (1/I)(\sum_{t=T}^{T+I}|ln(\hat{\varepsilon}_t^2) - ln(\hat{\sigma}_t^2)|)$$

Forecasting performance at different horizons

- *Mean Squared Error*

$$MSE = (1/I)(\sum_{t=T}^{T+I-\tau} (\hat{\varepsilon}^2_{t+\tau} -_\tau \hat{\sigma}^2_t)^2)$$

- *Mean Absolute Error.*

$$MAE = (1/I)(\sum_{t=T}^{T+I-\tau} |\hat{\varepsilon}^2_{t+\tau} -_\tau \hat{\sigma}^2_t|)$$

where ${}_\tau\hat{\sigma}^2_t$ is the forecast of the variance τ periods ahead given information at time t.

- *Mean Squared Error of the log of the squared residuals.*

$$[LE]^2 = (1/I)\left(\sum_{t=T}^{T+I-\tau} (\ln(\hat{\varepsilon}_{t+\tau}^2) - \ln(\hat{\sigma}_t^2))^2\right)$$

- *Mean Absolute Error of the log of the squared residuals.*

$$[MAE]^2 = (1/I)\left(\sum_{t=T}^{T+I-\tau} |\ln(\hat{\varepsilon}_{t+\tau}^2) - \ln(\hat{\sigma}_t^2)|\right)$$

where ${}_\tau\hat{\sigma}_t^2$ is the forecast of the variance τ periods ahead given information at time t.

Multivariate GARCH Models

- Multivariate extension of the GARCH(p,q) as follows:

$$y_t = \mu + u_t$$

$$vech(H_t) = C + \sum_{i=1}^{p} A_i vech(u_{t-i} u'_{t-i}) + \sum_{i=1}^{q} B_i vech(H_{t-i})$$

where $u_t | I_{t-1} \sim N(0, H_t)$

Multivariate GARCH Models

- $H_t = E(u_t u_t' | I_{t-1})$:$n \times n$ conditional variance matrix associated with u_t'
- $vech(H_t)$: $(n(n+1))/2 \times 1$ vector of all the unique elements of H_t
  - μ is $n \times 1$
  - C is $(n(n+1))/2 \times 1$
  - $A_1, A_2, ...A_p, B_1, ...B_q$ are $(n(n+1))/2 \times (n(n+1))/2$

Multivariate GARCH Models

Example

A bivariate GARCH (1,1)

$$\begin{bmatrix} x_t \\ w_t \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_w \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \nu_t \end{bmatrix}$$

Multivariate GARCH Models

Example

where

$$vech(H_t) = C + Avech(u_{t-1}u'_{t-1}) + Bvech(H_{t-1})$$

or

$$\begin{bmatrix} V_{t-1}(x_t) \\ COV_{t-1}(x_t w_t) \\ V_{t-1}(w_t) \end{bmatrix} = \begin{bmatrix} c_x \\ c_{xw} \\ c_w \end{bmatrix} + \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \begin{bmatrix} (\varepsilon_{t-1})^2 \\ (\varepsilon_{t-1}\nu_{t-1}) \\ (\nu_{t-1})^2 \end{bmatrix} + \begin{bmatrix} b_1 & b_2 & b_3 \\ b_4 & b_5 & b_6 \\ b_7 & b_8 & b_9 \end{bmatrix} \begin{bmatrix} V_{t-2}(x_{t-1}) \\ COV_{t-2}(x_{t-1} w_{t-1}) \\ V_{t-2}(w_{t-1}) \end{bmatrix}$$

A Diagonal Vech Parametrization

- Assume that each covariance depends only on its own past values and innovations
- Each element of H_t follows an univariate GARCH model driven by the corresponding cross product $u_t u_t'$.

$$\begin{bmatrix} c_x & c_{xw} \\ c_{xw} & c_w \end{bmatrix}, \begin{bmatrix} a_1 & a_5 \\ a_5 & a_9 \end{bmatrix}, \begin{bmatrix} b_1 & b_5 \\ b_5 & b_9 \end{bmatrix}$$

  are all positive definite.
- This can be ensured by doing simple Cholesky transformations to each of these matrices.

A Quadratic Specification

$$H_t = C'C + A\prime u_{t-1}u'_{t-1}A + B'H_{t-1}B.$$

C is a lower triangular with $n(n+1)/2$ parameters A and B are square matrices with n^2 parameters each

A Constant Correlation Specification

- Correlation between the assets is constant

$$COV_{t-1}(x_t w_t) = \rho \sqrt{V_{t-1}(x_t) V_{t-1}(w_t)}$$

- ρ is also estimated with the rest of the parameter set

Stationarity

- Conditions are similar to those discussed in the univariate case.
- The minimum square error forecast for $vech(H_t)$ ~$GARCH(1,1)$

$$E_s(vech(H_t)) = [\sum_{k=0}^{t-s-1}(A_1+B_1)^k]C + (A_1+B_1)^{t-s}vech(H_s)$$

- This can be derived in the usual way noting that

$$\begin{aligned} vech(H_t) &= C + Avech(u_{t-1}u_{t-1}') + Bvech(H_{t-1}), \\ vech(u_t u_t') &= vech(H_t) + vech(\eta_t) \\ vech(u_{t-1}u_{t-1}') &= vech(H_{t-1}) + vech(\eta_{t-1}), \end{aligned}$$

writing the stochastic representation

$$vech(u_t u_t') = C + (A+B)vech(u_{t-1}u_{t-1}') - Bvech(\eta_{t-1}) + vech(\eta_t),$$

and doing the same derivation as in the univariate case.

Stationarity

- Let $V\Lambda V^{-1}$ denote the Jordan decomposition of the matrix $A_1 + B_1$, so that

$$(A_1 + B_1)^{t-s} = V\Lambda^{t-s}V^{-1}$$

- $E_s(vech(H_t))$ converges to the vech of the unconditional covariance matrix, $vech(\Sigma) = (I - A_1 - B_1)^{-1}C \Leftrightarrow$ the absolute value of the largest eigen value of $A_1 + B_1$ is strictly less than one.
- If the absolute value of the largest eigen value of $A_1 + B_1$ is strictly less than one, then the forecast can be written

$$E_s(vech(H_t)) = vech(\Sigma) + (A_1 + B_1)^{t-s}(vech(H_s) - vech(\Sigma))$$

Co-Persistence in Variance

- The empirical estimates for univariate and multivariate ARCH models often indicate a high degree of persistence in the forecast moments of the conditional variances,

$$E_s(H_t)_{ii}, i = 1, 2, ...N, for\ t \to \infty.$$

- This persistence may be common across different series.

Definition

The multivariate ARCH process is said to be *co-persistent* in variance if at least one of the elements in $E_s(H_t)$ is non-convergent for increasing forecasts, $t - s$, but there exists a linear combination $\gamma'\varepsilon_t$, such that for every forecast origin s, the forecasts of the corresponding future conditional variances $E_s(\gamma' H_t \gamma)$ converge to a finite limit, independent of time s information.

Multivariate GARCH-M Models

Consider a system of n regression equations,

$$y_t = BX_t + Dvech(H_t) + u_t$$

$$vech(H_t) = C + \sum_{i=1}^{p} A_i vech(u_{t-i} u'_{t-i}) + \sum_{i=1}^{q} B_i vech(H_{t-i})$$

where

$$u_t | I_{t-1} \tilde{} N(0, H_t) B_{n \times k}, D_{n \times (n(n+1))/2}, C_{(n(n+1))/2 \times 1},$$
$$A_1, A_2, ...A_p, B_1, ...B_q are (n(n+1))/2 \times (n(n+1))/2$$

A Bivariate GARCH-M (1,1) Model

$$\begin{bmatrix} x_t \\ w_t \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_w \end{bmatrix} + \begin{bmatrix} d_1 & d_2 & d_3 \\ d_4 & d_5 & d_6 \end{bmatrix} \begin{bmatrix} V_{t-1}(x_t) \\ COV_{t-1}(x_t w_t) \\ V_{t-1}(w_t) \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \nu_t \end{bmatrix}$$

A Bivariate GARCH-M (1,1) Model

$$\begin{bmatrix} V_{t-1}(x_t) \\ COV_{t-1}(x_t w_t) \\ V_{t-1}(w_t) \end{bmatrix} = \begin{bmatrix} c_x \\ c_{xw} \\ c_w \end{bmatrix} + \begin{bmatrix} c_1 & c_2 & c_3 \\ c_4 & c_5 & c_6 \\ c_4 & c_5 & c_6 \end{bmatrix} \begin{bmatrix} (\varepsilon_{t-1})^2 \\ (\varepsilon_{t-1}\nu_{t-1}) \\ (\nu_{t-1})^2 \end{bmatrix} + \begin{bmatrix} b_1 & b_2 & b_3 \\ b_4 & b_5 & b_6 \\ b_4 & b_5 & b_6 \end{bmatrix} \begin{bmatrix} V_{t-2}(x_{t-1}) \\ COV_{t-2}(x_{t-1} w_{t-1}) \\ V_{t-2}(w_{t-1}) \end{bmatrix}$$

Estimation

- Conditional maximum likelihood estimation.

$$L_t(\theta) = -(n/2)\log(2\pi) - (1/2)\log(|H_t(\theta)|) - (1/2)u_t(\theta)'H_t^{-1}(\theta)u_t(\theta)$$

where θ represents a vector of parameters, n represents number of equations and t represents time.

Estimation

- Conditional on initial values for u_0 and H_0 , the likelihood function for the sample:

$$L(\theta) = \sum_{t=1}^{T} L_t(\theta)$$

- Model is highly non-linear and very unstable.

Testing the CAPM

- Problems:
  - The CAPM is a statement about the relationships between ex ante risk premiums and betas, both of which there are not directly observable.
  - The Roll- critique
  - The CAPM is a single-period model

The Unconditional CAPM

- In the simple version of the CAPM the expected return for asset j is

$$E(r_j) = r_f + \beta_{jm}(E(r_m) - r_f)$$

where r_j, r_f and r_m are the asset j, risk free and market rate of return respectively.

The Unconditional CAPM

Defining excess returns $\tilde{r}$, we can rewrite the above expressions as

$$E(\tilde{r}_j) = \beta_{jm}(E\tilde{r}_m)$$

where

$$\tilde{r}_j = r_j - r_f$$

$$\tilde{r}_m = r_m - r_f$$

The Conditional CAPM

- Models for r_{jt} in which the conditional density rather than the unconditional density returns is used.
- Conditional asset pricing model has

$$E(\tilde{r}_{jt}|I_{t-1}) = \beta_{jt} E(\tilde{r}_{mt}|I_{t-1})$$

where $\beta_{jt} = cov(\tilde{r}_{jt}\tilde{r}_{mt}|I_{t-1}) / var(\tilde{r}_{mt}|I_{t-1})$.

Multivariate GARCH-M Models: A CAPM with time varying covariances

- The conditional CAPM can be written as

$$E(r_{jt}|I_{t-1}) - r_{ft-1} = \beta_{jt}[E(r_{mt}|I_{t-1}) - r_{ft-1}]$$

where

$$\beta_{jt} = cov(r_{jt}r_{mt}|I_{t-1}) / var(r_{mt}|I_{t-1}).$$

and

$$H = \begin{bmatrix} var(r_{jt}|I_{t-1}) & cov(r_{jt}r_{mt}|I_{t-1}) \\ cov(r_{jt}r_{mt}|I_{t-1}) & var(r_{mt}|I_{t-1}) \end{bmatrix}$$

Multivariate GARCH-M Models: A CAPM with time varying covariances

- If we assume that the "market price of risk", λ is constant, where

$$\lambda = (E(r_{mt}|I_{t-1}) - r_{ft-1})/var(r_{mt}|I_{t-1})$$

- Then we can write

$$\begin{aligned} E(r_{jt}|I_{t-1}) - r_{ft-1} &= \beta_{jt}[E(r_{mt}|I_{t-1}) - r_{ft-1}] \\ &= \frac{cov(r_{jt}r_{mt}|I_{t-1})}{var(r_{mt}|I_{t-1})}[E(r_{mt}|I_{t-1}) - r_{ft-1}]. \\ &= \lambda cov(r_{jt}, r_{mt}|I_{t-1}) \end{aligned}$$

- And finally

$$r_{jt} = r_{ft-1} + \lambda cov(r_{jt}, r_{mt}|I_{t-1}) + u_{jt}$$

- Notice that

$$E(r_{mt}|I_{t-1}) - r_{ft-1} = \lambda var(r_{mt}|I_{t-1})$$

(since $\lambda = (E(r_{mt}|I_{t-1}) - r_{ft-1})/var(r_{mt}|I_{t-1})$)

$$r_{mt} = r_{ft-1} + \lambda var(r_{mt}|I_{t-1}) + u_{mt}$$

where u_{jt} and u_{mt} are the innovations.

- Time varying CAPM can be put into multivariate GARCH-M form as

$$y_t = b + dvech(H_t) + u_t$$

where $y_t = (r_{jt} - r_{ft-1}, r_{mt} - r_{ft-1})'$, $vech(H_t) = (var(r_{jt}|I_{t-1}),$ $cov(r_{jt}r_{mt}|I_{t-1}),$ $var(r_{mt}|I_{t-1}))'$, $u_t = (u_{jt}, u_{mt})'$ and

$$d = \lambda \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

- The zero restrictions implied by the theory may be tested by a likelihood ratio test which is asymptotically distributed $\chi^2(5)$.

Threshold Models

Series de Tiempo

july 2019

Introduction

- Economic variables may behave very differently in different states of the economy such as, for example, high/low inflation, high/low growth, or high/low stock prices (relative to dividends).
- A variety of nonlinear models have been proposed for describing the dynamics of economic time series subject to changes in regime [see, e.g., Tong (1983, 1990); Hamilton (1993); van Dijk et al. (2002); Dueker et al. (2007)].

TAR and SETAR Models

- Self-Exciting Threshold AutoRegressive (SETAR) models are an extension of autoregressive models which allow for a different behaviour of the series once the series enters a different regime.
- The switch from one regime to another depends on a past value of the series (hence the Self-Exciting portion of the name). crossing a threshold value. A possible parameterization of the SETAR model is

$$\begin{aligned} y_t &= y_{1t} && \text{if } y_{t-1} < k \\ &= y_{2t} && \text{if } y_{t-1} \geq k \end{aligned} \tag{1}$$

where

$$y_{it} = \mu_i + \sum_{j=1}^{p} \alpha_j^{(i)} y_{t-j} + \sigma_i u_t, \quad i = 1, 2, \tag{2}$$

- To estimate the model a grid has to be created for k, and then choose the value of the grid with highest likelihood. Notice that the model can be estimated by OLS and evaluated using MLE.

- In general we can consider a model such as

$$\begin{aligned} y_t &= y_{1t} && \text{if } z_t < k \\ &= y_{2t} && \text{if } z_t \geq k \end{aligned} \tag{3}$$

where z_t is a vector of exogenous or predetermined variables. When doing this we have to be careful with the exogeneity assumptions about z_t.

- To test for the existence of two states is not trivial because under the null that

$$\begin{aligned} \mu_1 &= \mu_2 \\ \alpha_j^{(1)} &= \alpha_j^{(2)} \text{ for } j = 1, ..., p. \end{aligned}$$

the parameter k is not defined, or is what is called a nuisance parameter.

Smooth Transition Autoregressive Models

- One posible drawback of the SETAR models is that potential big changes on the conditional mean process are trigerred by small changes in the variable that drives the regimes.
- The STAR model may be thought of as a function of two (or more) autoregressive processes which are averaged, at any given point in time, according to some continuous function $G(\cdot)$ taking values in $[0, 1]$.
- STAR models for the univariate time series $\{x_t\}$ may be formulated as

$$y_t = G(\mathbf{z}_{t-1})y_{1t} + [1 - G(\mathbf{z}_{t-1})]y_{2t}, \quad t = 1, 2, \ldots, \tag{4}$$

where $\mathbf{z}_{t-1}$ is a vector of exogenous and/or pre-determined variables and

$$y_{it} = \mu_i + \sum_{j=1}^{p} \alpha_j^{(i)} y_{t-j} + \sigma_i u_t, \quad i = 1, 2, \tag{5}$$

where p is a positive integer, $\{u_t\}$ are independent and identically distributed (i.i.d.) random variables such that u_t is independent of $(y_{t-p}, \ldots, y_0)$ $\mathbb{E}(u_t) = \mathbb{E}(u_t^2 - 1) = 0$, σ_1 and σ_2 are positive constants, and μ_i and $\alpha_j^{(i)}$ $(i = 1, 2;\ j = 1, \ldots, p)$ are real constants.

Logistic Smooth Transition Autoregressive Models

- STAR models like (4)–(5) have been used extensively in the analysis of economic and financial data. The main feature that differentiates alternative STAR specifications is the choice of the mixing (or transition) function $G(\cdot)$ and the transition variables $\mathbf{z}_{t-1}$ (see Teräsvirta, 1998; van Dijk et al., 2002). A popular choice for $G(\cdot)$ in (4) is the logistic specification

$$G(y_{t-1}) = \frac{\exp\{-\gamma(y_{t-1} - y^*)\}}{1 + \exp\{-\gamma(y_{t-1} - y^*)\}}, \qquad \gamma > 0, \tag{6}$$

  which gives rise to the pth-order LSTAR model, or LSTAR(p).
- The location parameter y^* in (6) may be interpreted as the threshold between the two regimes associated with the extreme values $\lim_{y_{t-1}\to\infty} G(y_{t-1}) = 0$ and $\lim_{y_{t-1}\to-\infty} G(y_{t-1}) = 1$, while the slope parameter γ determines the smoothness of the transitions between the two regimes.

Exponential Smooth Transition Autoregressive Models

- ESTAR models are intended to capture a different behavior of the time series when is close to the threshold from when it is far from the threshold.

$$G(y_{t-1}) = \exp\{-\gamma(y_{t-1} - y^*)^2\}, \qquad \gamma > 0, \tag{7}$$

  which gives rise to the pth-order ESTAR model, or ESTAR(p). Notice that the extreme values $\lim_{y_{t-1}\to\infty} G(y_{t-1}) = 0$ and $\lim_{y_{t-1}\to-\infty} G(y_{t-1}) = 0$, while $\lim_{y_{t-1}\to y^*} G(y_{t-1}) = 1$.
- This type of models are used for situations like target zones where the behaviour of the exchange rate in the middle of the band is different from that in the middle of the band.

Univariate Contemporaneous-Threshold Models

- The C-STAR model of Dueker et al. (2007) may be thought of as a function of two (or more) autoregressive processes which are averaged, at any given point in time, according to some continuous function $G(\cdot)$ taking values in $[0, 1]$.
- STAR models for the univariate time series $\{y_t\}$ may be formulated as

$$y_t = G(\mathbf{z}_{t-1})y_{1t} + [1 - G(\mathbf{z}_{t-1})]y_{2t}, \quad t = 1, 2, \ldots,$$

where $\mathbf{z}_{t-1}$ is a vector of exogenous and/or pre-determined variables and

$$y_{it} = \mu_i + \sum_{j=1}^{p} \alpha_j^{(i)} y_{t-j} + \sigma_i u_t, \quad i = 1, 2,$$

where p is a positive integer, $\{u_t\}$ are independent and identically distributed (i.i.d.) random variables such that u_t is independent of $(y_{t-p}, \ldots, y_0)$, $\mathbb{E}(u_t) = \mathbb{E}(u_t^2 - 1) = 0$, σ_1 and σ_2 are positive constants, and μ_i and $\alpha_j^{(i)}$ $(i = 1, 2; j = 1, \ldots, p)$ are real constants.

- The feature that differentiates alternative STAR models is the choice of the mixing function $G(\cdot)$ and transition variables $\mathbf{z}_{t-1}$ [cf. Teräsvirta (1998); van Dijk et al. (2002)].
- The C-STAR model of order p is obtained by defining the mixing function $G(\cdot)$ in (4) as

$$G(\mathbf{z}_{t-1}) = \frac{\mathbb{P}(y_{1t} < y^* | \mathbf{z}_{t-1}; \boldsymbol{\vartheta}_1)}{\mathbb{P}(y_{1t} < y^* | \mathbf{z}_{t-1}; \boldsymbol{\vartheta}_1) + \mathbb{P}(y_{2t} \geqslant y^* | \mathbf{z}_{t-1}; \boldsymbol{\vartheta}_2)}$$

where $\boldsymbol{\vartheta}_i = (\mu_i, \alpha_1^{(i)}, \ldots, \alpha_p^{(i)}, \sigma_i^2)'$ is the vector of parameters associated with regime i, and $\mathbf{z}_{t-1} = (y_{t-1}, ..., y_{t-p})'$.

- Letting $\boldsymbol{\alpha}_i = (\alpha_1^{(i)}, \ldots, \alpha_p^{(i)})'$ $(i = 1, 2)$, the (conditionally) Gaussian two-regime

$$G(\mathbf{z}_{t-1}) = \frac{\Phi(\frac{y^* - \mu_1 - \boldsymbol{\alpha}_1' \mathbf{z}_{t-1}}{\sigma_1})}{\Phi(\frac{y^* - \mu_1 - \boldsymbol{\alpha}_1' \mathbf{z}_{t-1}}{\sigma_1}) + 1 - \Phi(\frac{y^* - \mu_2 - \boldsymbol{\alpha}_2' \mathbf{z}_{t-1}}{\sigma_2})}, \qquad (8)$$

where $\Phi(\cdot)$ denotes the standard normal distribution function and y^* is a threshold parameter.

- Hence, (4) may be rewritten as

$$y_t = \frac{\mathbb{P}(y_{1t} < y^* | \mathbf{z}_{t-1}; \boldsymbol{\vartheta}_1) y_{1t} + \mathbb{P}(y_{2t} \geq y^* | \mathbf{z}_{t-1}; \boldsymbol{\vartheta}_2) y_{2t}}{\mathbb{P}(y_{1t} < y^* | \mathbf{z}_{t-1}; \boldsymbol{\vartheta}_1) + \mathbb{P}(y_{2t} \geq y^* | \mathbf{z}_{t-1}; \boldsymbol{\vartheta}_2)}.$$

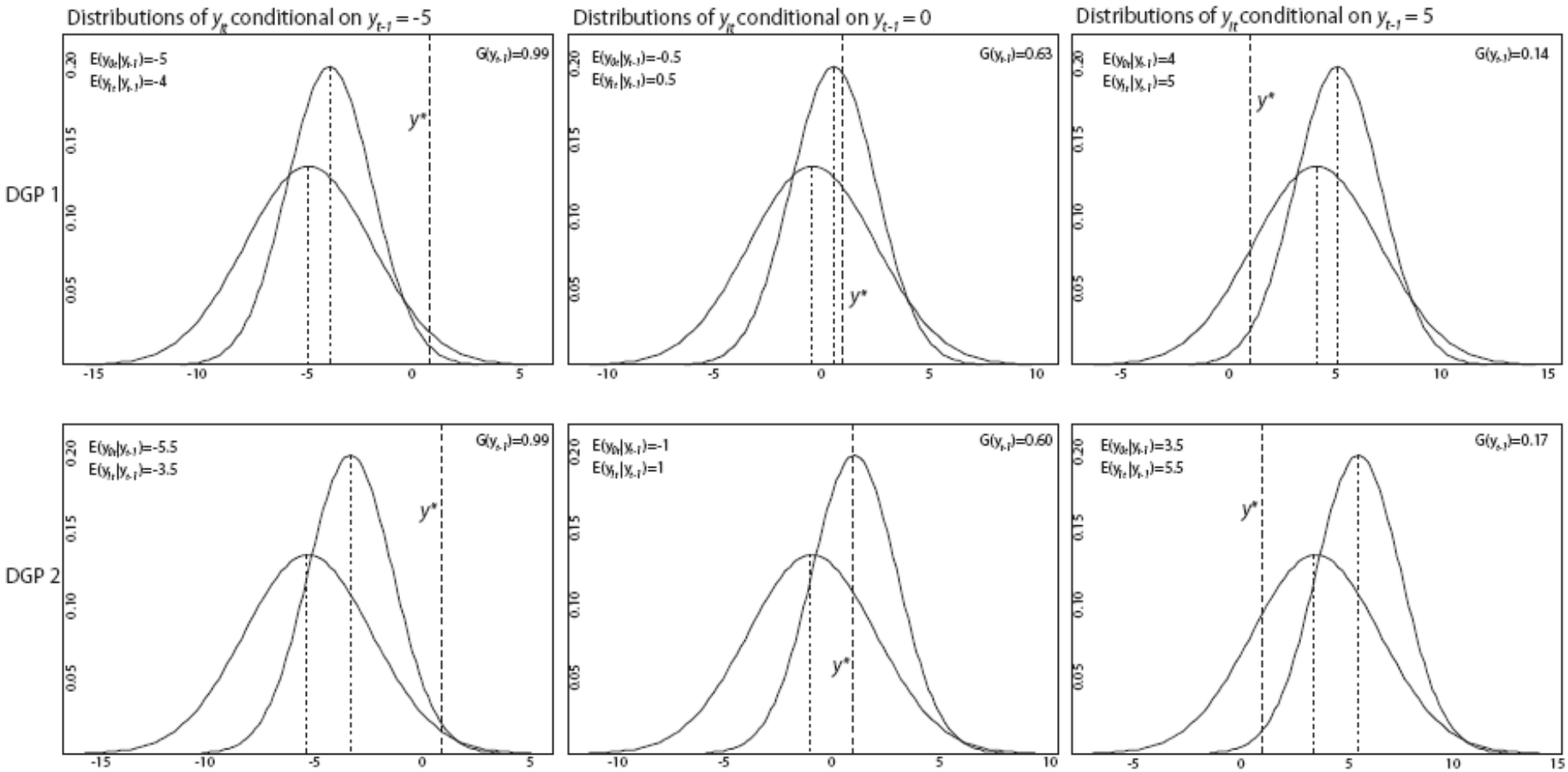
- As with conventional STAR models, a C-STAR model may be thought of as a regime-switching model that allows for two regimes associated with the two latent variables y_{1t} and y_{2t}.
- Alternatively, a C-STAR model may be thought of as allowing for a continuum of regimes, each of which is associated with a different value of $G(\mathbf{z}_{t-1})$.

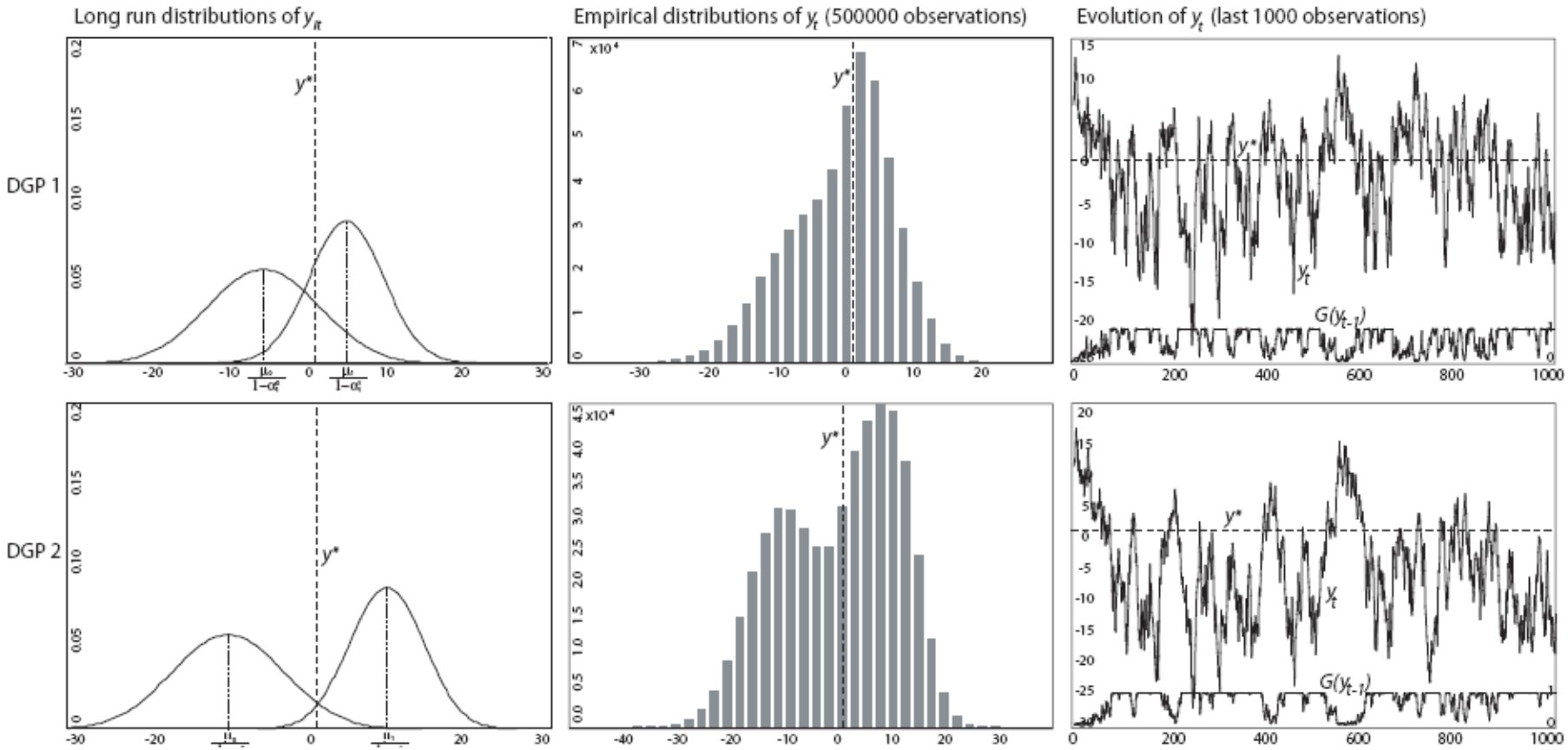
- One of the main purposes of the C-STAR model is to address two somewhat arbitrary features of conventional STAR models:
- First, STAR models specify a delay such that the mixing function for period t consists of a function of y_{t-j} for some $j \geq 1$.
- Second, STAR models specify which of and in what way the model parameters enter the mixing function.
- C-STAR models address these twin issues in an intuitive way: they use a forecasting function such that the mixing function depends on the ex ante regime-dependent probabilities that y_t will exceed the threshold value(s). Furthermore, the mixing function makes use of all of the model parameters in a coherent way.

Table 2. DGPs

	μ_0	α_1^0	σ_0	μ_1	α_1^1	σ_1	y^*
DGP 1	-0.5	0.9	3	0.5	0.9	2	1
DGP 2	-1	0.9	3	1	0.9	2	1

Conditional Distributions




Long run distributions of y_{it}
Empirical distributions of y_t (500000 observations)
Evolution of y_t (last 1000 observations)
DGP 1
DGP 2
y*
y_t
$G(y_{t-1})$
x10⁴

Stability Properties of the Skeleton of the C-STAR

- As Chan and Tong (1985) pointed out, we can analyze the properties of a nonlinear time series by considering the deterministic part of the model alone.
- This part is usually called the skeleton of the model and is defined as $y_t = F(y_{t-1}, \Theta)$, where

$$F(y_{t-1}, \Theta) = G(y_{t-1})(\mu_0 + \alpha_1^0 y_{t-1}) + (1 - G(y_{t-1}))(\mu_1 + \alpha_1^1 y_{t-1}), \quad (9)$$

  and $\Theta = \{\Theta_0, \Theta_1, y^*\}$.
- Then a fixed point of the skeleton of the model is any value, y_L, that satisfies

$$y_L = F(y_L, \Theta) = G(y_L)(\mu_0 + \alpha_1^0 y_L) + (1 - G(y_L))(\mu_1 + \alpha_1^1 y_L). \quad (10)$$

- Since the C-STAR(1) is a nonlinear model, there may be one, several or no equilibrium values that satisfy equation (10).
- Then, assessing which of the equilibria of the nonlinear first-order difference equation are stable is crucial for learning about the stability properties of the C-STAR model.
- We use each of the DGPs presented in Table 2 to assess:
- *i*) the number of equilibria and
- *ii*) the stability of the equilibria.

- We find the number of equilibria for the different DGPs presented in Table 1, using a grid of starting values to solve equation (10) numerically.
- For each equilibrium, we analyze whether it is locally stable by considering the following expansion around the fix point

$$\begin{aligned} y_t - y_L &= F(y_{t-1}, \Theta) - F(y_L, \Theta) \\ &\simeq \frac{\partial F(y_{t-1}, \Theta)}{\partial y_{t-1}} (y_{t-1} - y_L). \end{aligned} \tag{11}$$

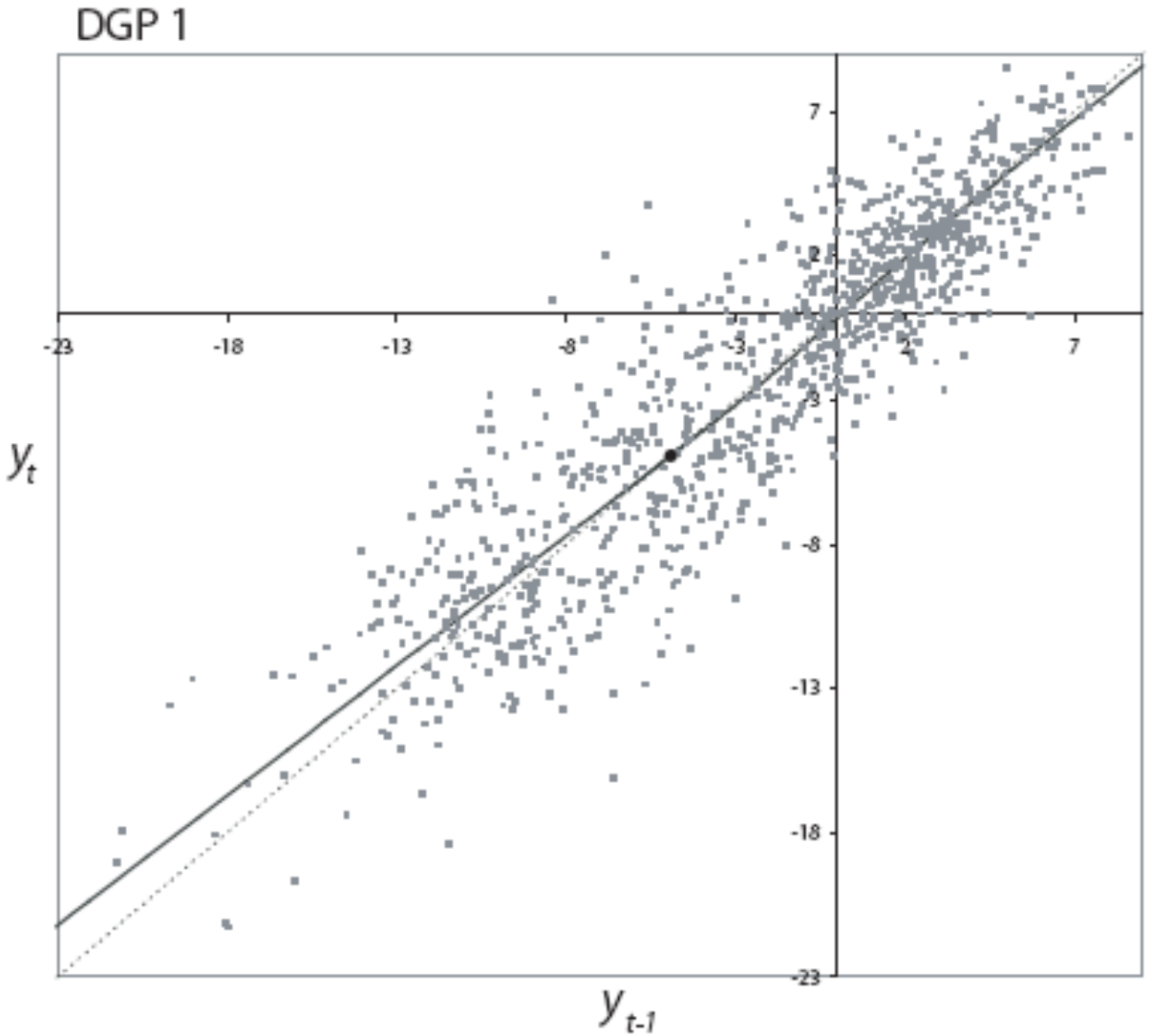
- Whenever $\left| \frac{\partial F(y_{t-1}, \Theta)}{\partial y_{t-1}} \right| < 1$, the equilibrium is locally stable and $F(y_{t-1}, \Theta)$ is a contraction in the neighborhood of $y = y_L$.

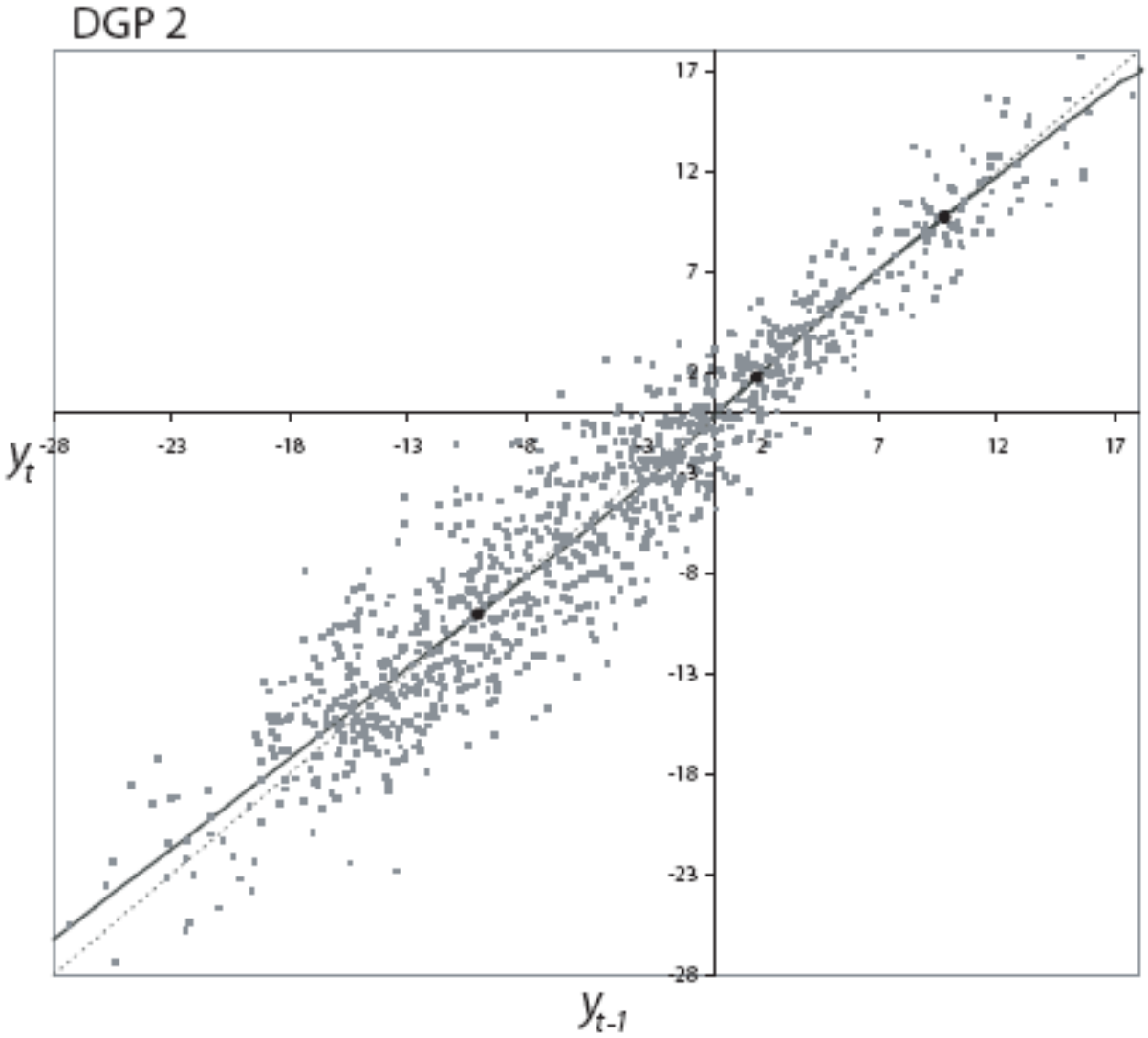
- Where

$$\frac{\partial F(y_L,\Theta)}{\partial y_{t-1}} = \alpha_1^1 + (\alpha_1^0 - \alpha_1^1)G(y_L) + \left[(\mu_0 - \mu_1) + (\alpha_1^0 - \alpha_1^1)y_L\right]\frac{\partial G(y_L)}{\partial y_{t-1}} \tag{12}$$

and

$$\begin{aligned}\frac{\partial G(y_L)}{\partial y_{t-1}} &= \frac{-\left(\frac{\alpha^0}{\sigma_0}\phi(w_0^L)\Phi(w_1^L) + \frac{\alpha^1}{\sigma_1}\phi(w_1^L)\Phi(w_0^L)\right)}{\left(\Phi(w_0^L) + [1-\Phi(w_1^L)]\right)^2}, \text{ where } \phi = \Phi', \\ w_0^L &= (y^* - \mu_0 - \alpha_1^0 y_L)/\sigma_0 \text{ and } w_1^L = (y^* - \mu_1 - \alpha_1^1 y_L)/\sigma_1.\end{aligned}$$


DGP 1
7
2
-23
-18
-13
-8
-3
2
7
-3
-8
-13
-18
-23
y_t
y_{t-1}


DGP 2
17
12
7
2
-3
-8
-13
-18
-23
-28
-28
-23
-18
-13
-8
-3
2
7
12
17
y_t
y_{t-1}

State Dependent Contemporaneous-Threshold Models

- Here we investigate the possibility that the separation of regimes implied by nonlinear regime-switching autoregressive models is better characterized in relative terms rather than being dictated by the (constant) level of some variables.
- We consider the threshold as function of variables which potentially affect the evolution of the time series under consideration.
- For example, interest rates may be considered high or low not in absolute terms but relative to relevant macroeconomic variables that describe the state of the economy. A rate of interest which is considered to be high when the economy is in a state of low inflation and high output growth may be considered too low when the economy in a state of high inflation and low output growth.
- We propose to model such behavior using variants of smooth transition autoregressive (STAR) models which allow the interest rate threshold to evolve over time as a function of inflation and output growth.

State Dependent Contemporaneous-Threshold Models

- We generalize the standard C-STAR specifications to allow for state dependency of the threshold.
- We define y^*_{t-1} as a time-varying threshold. This is specified as a linear combination of the elements of a k-dimensional vector $\mathbf{x}_{t-1} = (x_{1,t-1}, \ldots, x_{k,t-1})'$ of observable exogenous and/or predetermined variables, that is,

$$y^*_{t-1} = y^* + \boldsymbol{\delta}' \mathbf{x}_{t-1}, \tag{13}$$

where y^* is an unknown threshold intercept and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_k)'$ is a vector of unknown parameters.

State Dependent Contemporaneous-Threshold Models

- The SDC-STAR(p) model
- To generalize the C-STAR(p) specification to allow for a state-dependent threshold, the mixing weights in (8) are replaced with

$$G(\mathbf{z}_{t-1}) = \frac{F_1\left(\frac{y^*_{t-1} - \mu_1 - \boldsymbol{\alpha}_1'\mathbf{y}_{t-1}}{\sigma_1}\right)}{F_1\left(\frac{y^*_{t-1} - \mu_1 - \boldsymbol{\alpha}_1'\mathbf{y}_{t-1}}{\sigma_1}\right) + 1 - F_2\left(\frac{y^*_{t-1} - \mu_2 - \boldsymbol{\alpha}_2'\mathbf{y}_{t-1}}{\sigma_2}\right)}, \tag{14}$$

where y^*_{t-1} is a time-varying threshold depending on the observable exogenous and/or predetermined variables $\mathbf{x}_{t-1}$, and $\mathbf{z}_{t-1} = (\mathbf{y}'_{t-1}, \mathbf{x}'_{t-1})'$. As before, the threshold y^*_{t-1} is specified to be the linear combination of the elements of $\mathbf{x}_{t-1}$ given in (13).

State Dependent Contemporaneous-Threshold Models

- Notice that

$$G(\mathbf{z}_{t-1}) = \frac{\mathbb{P}(y_{1t} < y^*_{t-1}|\mathbf{z}_{t-1})}{\mathbb{P}(y_{1t} < y^*_{t-1}|\mathbf{z}_{t-1}) + \mathbb{P}(y_{2t} \geq y^*_{t-1}|\mathbf{z}_{t-1})}$$

and

$$1 - G(\mathbf{z}_{t-1}) = \frac{\mathbb{P}(y_{2t} \geq y^*_{t-1}|\mathbf{z}_{t-1})}{\mathbb{P}(y_{1t} < y^*_{t-1}|\mathbf{z}_{t-1}) + \mathbb{P}(y_{2t} \geq y^*_{t-1}|\mathbf{z}_{t-1})}.$$

- Hence, under the assumptions in (13)–(14), the SDC-STAR(p) model may be rewritten as

$$y_t = \frac{\mathbb{P}(y_{1t} < y^*_{t-1}|\mathbf{z}_{t-1})y_{1t} + \mathbb{P}(y_{2t} \geq y^*_{t-1}|\mathbf{z}_{t-1})y_{2t}}{\mathbb{P}(y_{1t} < y^*_{t-1}|\mathbf{z}_{t-1}) + \mathbb{P}(y_{2t} \geq y^*_{t-1}|\mathbf{z}_{t-1})}. \tag{15}$$

  As in the case of the C-STAR, the mixing weights involve the probability that the contemporaneous value of y_{1t} (y_{2t}) is smaller (greater) than some threshold level y^*_{t-1}.
- The SDC-STAR(p) model reduces to a C-STAR(p) under the restriction $\delta = \mathbf{0}$.

State Dependent Contemporaneous-Threshold Models

- A SDC-STARX(p) Model.
- This model includes exogenous and/or predetermined variables $\mathbf{x}_{t-1}$ in the equation that describes the dynamics of the latent variables y_{1t} and y_{2t}. More specifically, let

$$y_{it} = \mu_i + \boldsymbol{\alpha}^{(i)\prime}\mathbf{y}_{t-1} + \boldsymbol{\delta}^{(i)\prime}\mathbf{x}_{t-1} + \sigma_i u_{it}, \qquad i = 1, 2, \tag{16}$$

where $\boldsymbol{\delta}^{(i)} = (\delta_1^{(i)}, \ldots, \delta_k^{(i)})'$ are unknown parameters, and take $G(\cdot)$ to have the form

$$G(\mathbf{z}_{t-1}) = \frac{F_1\left(\dfrac{y^*_{t-1} - \mu_1 - \boldsymbol{\beta}_1'\mathbf{z}_{t-1}}{\sigma_1}\right)}{F_1\left(\dfrac{y^*_{t-1} - \mu_1 - \boldsymbol{\beta}_1'\mathbf{z}_{t-1}}{\sigma_1}\right) + 1 - F_2\left(\dfrac{y^*_{t-1} - \mu_2 - \boldsymbol{\beta}_2'\mathbf{z}_{t-1}}{\sigma_2}\right)}, \tag{17}$$

with $\boldsymbol{\beta}_i = (\boldsymbol{\alpha}_i', \boldsymbol{\delta}_i')'$ $(i = 1, 2)$. Equations (4), (13), (16) and (17) define a pth-order state-dependent C-STAR model with exogenous variables.

State Dependent Contemporaneous-Threshold Models

- The SDC-STARX(p) model nests several specifications.
- If $\delta^{(1)} = \delta^{(2)}$, $\delta = \mathbf{0}$, we obtain a restricted SDC-STARX(p) model, or *RSDC-STARX*(p), the mixing weights of which are the same as those of a SDC-STAR(p) model with threshold $y^*_{t-1} = y^* - \delta^{(1)\prime}\mathbf{x}_{t-1}$; the two specifications are not, however, equivalent because they imply different conditional distributions for the latent variables y_{it}.
- If $\delta^{(1)} = \delta^{(2)} = \mathbf{0}$, the SDC-STARX($p$) model becomes a SDC-STAR(p).
- If $\delta = \mathbf{0}$, we obtain a pth-order C-STAR model with exogenous variables, or C-STARX(p), and constant threshold y^*.
- Finally, if $\delta^{(1)} = \delta^{(2)} = \delta = \mathbf{0}$, the SDC-STARX($p$) reduces to a C-STAR($p$) with constant threshold y^*.
- The validity of these restrictions on $\delta^{(1)}$, $\delta^{(2)}$ and δ can be tested using, for example, conventional likelihood ratio (LR) tests.

Stability

- The stability properties of SDC-STAR and C-STARX models follow as special cases of the SDC-STARX.
- Consider a SDC-STARX(1) model with a threshold which depends on a single exogenous variable x_{t-1} $(k = 1)$. The skeleton of such a model is defined as

$$Y_t = S(Y_{t-1}, X_{t-1}), \tag{18}$$

where

$$\begin{aligned} S(Y_{t-1}, X_{t-1}) &= G(Y_{t-1}, X_{t-1})\{\mu_1 + \alpha_1^{(1)} Y_{t-1} + \delta_1^{(1)} X_{t-1}\} \\ &\quad + \{1 - G(Y_{t-1}, X_{t-1})\}\{\mu_2 + \alpha_1^{(2)} Y_{t-1} + \delta_1^{(2)} X_{t-1}\}, \end{aligned} \tag{19}$$

and $G(Y_{t-1}, X_{t-1})$ is given by (17) with $\mathbf{z}_{t-1} = (Y_{t-1}, X_{t-1})'$ and $y^*_{t-1} = y^* + \delta x_{t-1}$.

Stability

- Assuming $\{x_t\}$ is stationary, a fixed point of the skeleton is any value Y_e which satisfies the equation

$$Y_e = S(Y_e, X_e), \tag{20}$$

where $X_e = \mathbb{E}(x_t)$. The value Y_e is said to be an equilibrium point of the SDC-STARX(1) model and, since the model is nonlinear, there may be one, several or no equilibrium points satisfying (20).

- An examination of the local stability of each of the equilibrium points can be carried out by considering a first-order Taylor expansion about the fixed point,

$$Y_t - Y_e \approx \lambda(Y_e, X_e)(Y_{t-1} - Y_e), \tag{21}$$

where

$$\lambda(Y_e, X_e) = \left. \frac{\partial S(Y_{t-1}, X_e)}{\partial Y_{t-1}} \right|_{Y_{t-1}=Y_e}. \tag{22}$$

If $|\lambda(Y_e, X_e)| < 1$, then the equilibrium is locally stable and Y_t is a contraction in the neighbourhood of (Y_e, X_e).

Stability

- It is straightforward to verify that

$$\frac{\partial S(Y_{t-1}, X_e)}{\partial Y_{t-1}} = \alpha_1^{(2)} + \{\alpha_1^{(1)} - \alpha_1^{(2)}\} G(Y_{t-1}, X_e)$$
$$+ \left(\mu_1 - \mu_2 + \{\alpha_1^{(1)} - \alpha_1^{(2)}\} Y_{t-1} + \{\delta_1^{(1)} - \delta_1^{(2)}\} X_e\right) \frac{\partial G(Y_{t-1}, X_e)}{\partial Y_{t-1}},$$

where

$$\frac{\partial G(Y_{t-1}, X_e)}{\partial Y_{t-1}} = -\frac{\sigma_1^{-1}\alpha_1^{(1)} f_1(w_1)[1 - F_2(w_2)] + \sigma_2^{-1}\alpha_1^{(2)} f_2(w_2) F_1(w_1)}{\{F_1(w_1) + 1 - F_2(w_2)\}^2}, \tag{23}$$

$f_i(\cdot)$ is the density of $F_i(\cdot)$ $(i = 1, 2)$, and

$$w_i = \sigma_i^{-1}\left(y^* - \mu_i - \alpha_1^{(i)} Y_{t-1} + \{\delta - \delta_1^{(i)}\} X_e\right), \qquad i = 1, 2.$$

- A numerical illustration
- Consider a SDC-STARX(1) model where u_{it} $(i = 1, 2)$ have the Student-t distribution with ν_i degrees of freedom. Assume that $\{x_t\}$ is the autoregressive model

$$x_t = 0.2 + 0.9x_{t-1} + \varepsilon_t,$$

  where $\{\varepsilon_t\}$ are i.i.d. random variables, independent of $\{u_{1t}\}$ and $\{u_{2t}\}$, having the standardized Student-t distribution with 5 degrees of freedom.
- The parameters of the model take the following values:

$$\mu_1 = -0.5, \quad \mu_2 = 0.5, \quad \alpha_1^{(1)} = 0.9, \quad \alpha_1^{(2)} = 0.9, 0.99,$$
$$\sigma_1 = 3, \quad \sigma_2 = 2, \nu_1 = 3, \quad \nu_2 = 4, \quad y^* = 0,$$
$$\delta = 0, 0.3, \quad \delta_1^{(1)} = 0, 0.1, \quad \delta_1^{(2)} = 0, -0.1.$$

  For each parameter configuration, we use a grid of starting values to solve equation (20) numerically and find the equilibrium points; the local stability of each equilibrium point is then examined by considering the expansion in (21)–(23).

Stability

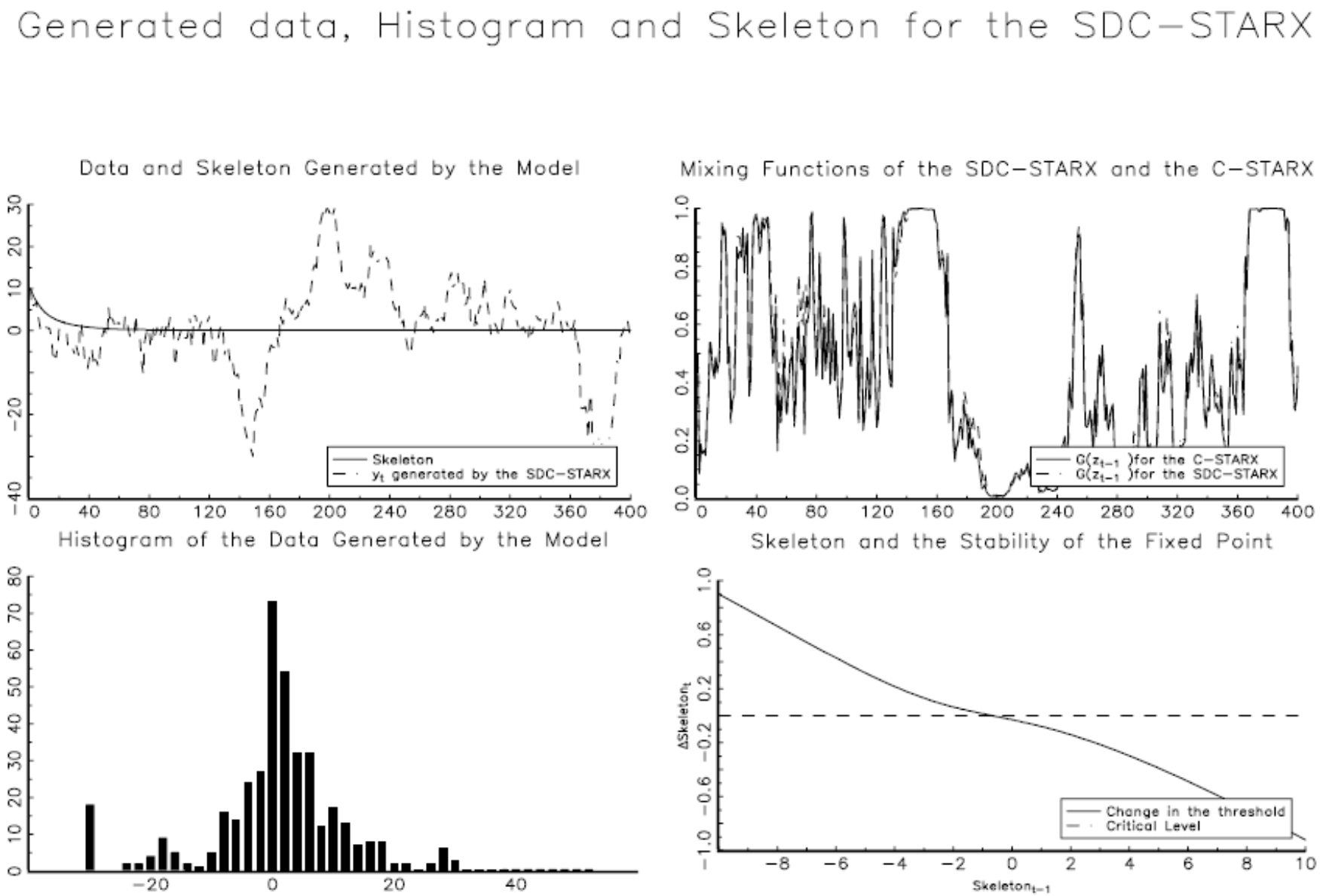


Figure 1

Stability

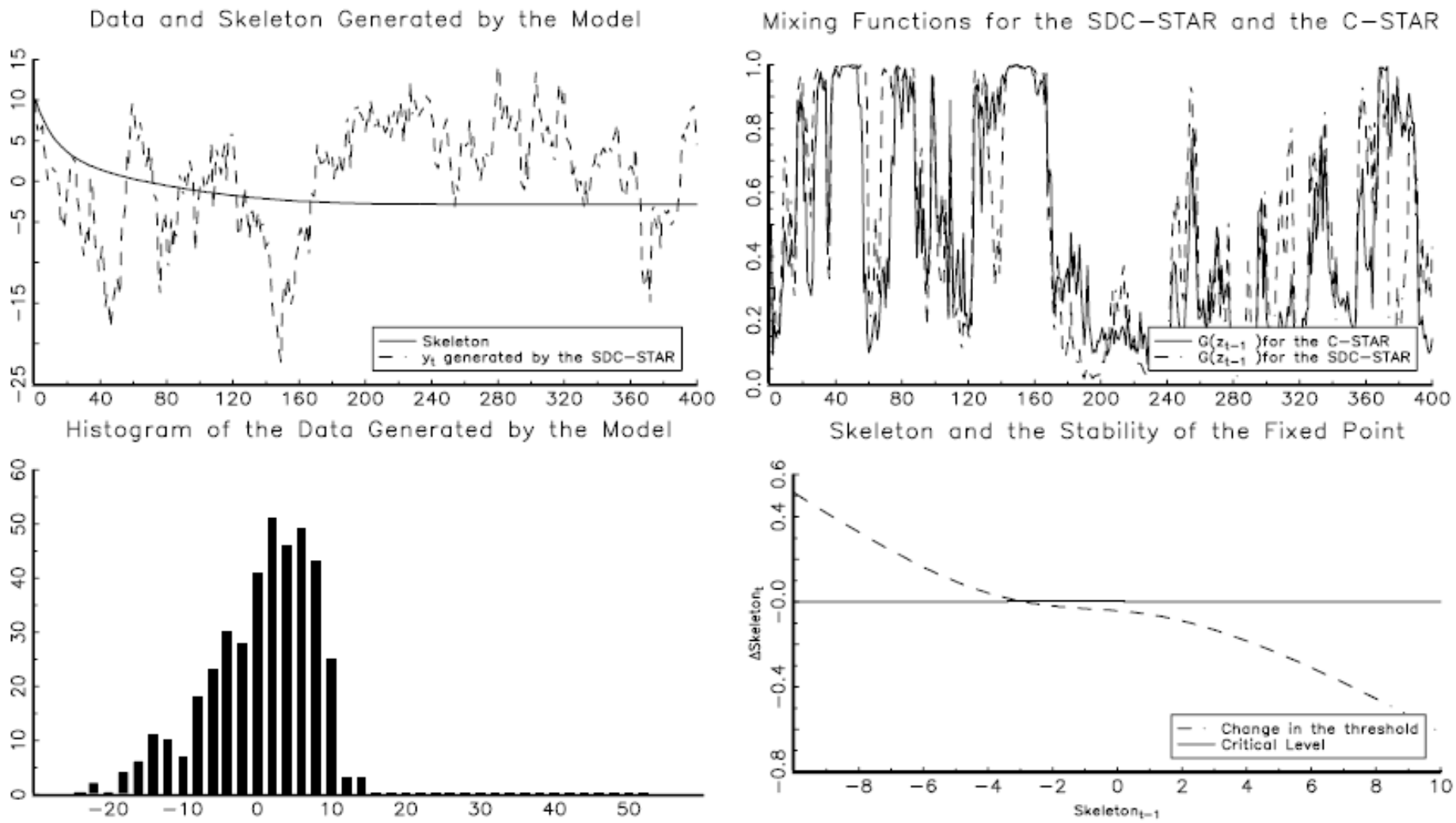


Figure 2

Estimation for the SDC-STAR model

- Given u_{1t} and u_{2t}, the parameters can be estimated by the ML method.
- Letting $\boldsymbol{\theta}$ denote the vector of all the unknown parameters, the conditional log-likelihood function associated with a sample $(y_1, \ldots, y_T)$ from the is

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=p+1}^{T} \ln \ell_t(\boldsymbol{\theta}),$$

where

$$\begin{aligned} \ell_t(\boldsymbol{\theta}) \quad = \quad & \frac{G(\mathbf{z}_{t-1})}{\sigma_1} f_1\left(\frac{y_t - \mu_1 - \boldsymbol{\alpha}_1' \mathbf{y}_{t-1}}{\sigma_1}\right) \\ & + \frac{1 - G(\mathbf{z}_{t-1})}{\sigma_2} f_2\left(\frac{y_t - \mu_2 - \boldsymbol{\alpha}_2' \mathbf{y}_{t-1}}{\sigma_2}\right), \end{aligned} \tag{24}$$

with $G(\mathbf{z}_{t-1})$ given by (14) and $f_i(\cdot)$ denoting, as before, the probability density function of u_{it} $(i = 1, 2)$.

Estimation for the SDC-STARX

- Under this model, the contribution of the t-th observation to the conditional likelihood is

$$\begin{aligned}\ell_t(\boldsymbol{\theta}) &= \frac{G(\mathbf{z}_{t-1})}{\sigma_1} f_1\left(\frac{y_t - \mu_1 - \boldsymbol{\beta}_1' \mathbf{z}_{t-1}}{\sigma_1}\right) \\ &\quad + \frac{1 - G(\mathbf{z}_{t-1})}{\sigma_2} f_2\left(\frac{y_t - \mu_2 - \boldsymbol{\beta}_2' \mathbf{z}_{t-1}}{\sigma_2}\right),\end{aligned}$$

  with $G(\mathbf{z}_{t-1})$ given by (17).

- In the remainder of the paper, we use the standardized Student-t distribution with $\nu_i > 2$ degrees of freedom, i.e.,

$$f_i(z) = \frac{\Gamma\left(\{\nu_i + 1\}/2\right)}{\Gamma(\nu_i/2)\sqrt{(\nu_i - 2)\pi}} \left(1 + \frac{z^2}{\nu_i - 2}\right)^{-(\nu_i+1)/2}, \qquad -\infty < z < \infty,$$

  where $\Gamma(\cdot)$ is the gamma function. This specification has greater flexibility to accommodate possible outliers and other heavy-tailed characteristics in the data.

An Empirical Application

- We investigate whether the dynamics of U.S. short-term interests can be adequately described using C-STAR model with constant threshold or whether they are better represented by the SDC-STAR and SDC-STARX models with a threshold which is determined by inflation and output growth.
- Our data set consists of quarterly observations from 1947:2 to 2008:4 on the secondary market rate on three-month U.S. Treasury bills (y_t), the growth rate of real gross domestic product (g_t), and the consumer price inflation rate (π_t).

Estimation Results

- We first investigate the presence of nonlinearities in the dynamics of the interest rate. Specifically, using the modified Hansen LR-based method, we test a linear AR(4) model for the interest rate against C-STAR(4) alternative.
- The results of the tests show that the standardized LR statistic has, for all choices of the bandwidth parameter M, a P-value smaller than 0.05 (0.10) for the C-STAR(4) model. In view of the fact that the test is conservative by construction, these results provide significant evidence in favour of the nonlinear models.
- The state-dependent threshold y_{t-1}^* is specified as in (13) with $\mathbf{x}_{t-1} = (g_{t-1}, \pi_{t-1})'$.

Table 4: C-STAR, SDC-STAR and C-STARX Models

	C-STAR(4)	SDC-STAR(4)	RSDC-STARX(4)	C-STARX(4)	SDC-STARX(4)
μ_1	0.0491 (0.0503)	0.1440 (0.0479)	−0.0465 (0.0603)	0.0063 (0.0558)	−0.0607 (0.0513)
μ_2	−0.1046 (0.3341)	−0.0381 (0.2771)	−0.2268. (0.3734)	−0.5744 (0.2242)	−0.8349 (0.2893)
$\alpha_1^{(1)}$	1.3072 (0.0812)	1.2230 (0.0634)	1.3080 (0.1000)	1.3003 (0.0718)	1.2706 (0.0644)
$\alpha_2^{(1)}$	−0.3042 (0.1313)	−0.2114 (0.0857)	−0.3358 (0.1615)	−0.2928 (0.1163)	−0.2508 (0.0792)
$\alpha_3^{(1)}$	0.1142 (0.1035)	0.1102 (0.0846)	0.1152 (0.1108)	0.0980 (0.0956)	0.0537 (0.0783)
$\alpha_4^{(1)}$	−0.1023 (0.0635)	−0.0859 (0.0565)	−0.0708 (0.0650)	−0.0889 (0.0590)	−0.0397 (0.0531)
$\alpha_1^{(2)}$	1.0726 (0.1248)	1.2128 (0.1288)	1.0613. (0.1309)	0.9639 (0.09483)	0.9949 (0.1055)
$\alpha_2^{(2)}$	−0.0771 (0.1885)	−0.0989 (0.1719)	−0.0657 (0.1932)	−0.0773 (0.1458)	−0.0319 (0.1501)
$\alpha_3^{(2)}$	0.2700 (0.1604)	0.2859 (0.1445)	0.2969 (0.1693)	0.4162 (0.1320)	0.4729 (0.1376)
$\alpha_4^{(2)}$	−0.2693 (0.1053)	−0.4070. (0.0818)	−0.2940 (0.1062)	−0.3273 (0.0838)	−0.4510 (0.0728)

σ_1	0.3078 (0.0924)	0.3998 (0.1215)	0.3100 (0.0874)	0.2303 (0.0481)	0.4458 (0.1269)
σ_2	1.2893 (0.2901)	1.2189 (0.2445)	1.2648 (0.2877)	1.5630 (0.4293)	0.8953 (0.1317)
y^*	5.4196 (0.3286)	4.7219 (1.2844)	5.5435 (0.3303)	5.3874 (0.2720)	5.0555 (0.9534)
$\delta_1^{(1)}$	—	—	0.0156 (0.0062)	0.0054 (0.0051)	0.0102 (0.0049)
$\delta_2^{(1)}$	—	—	−0.0026 (0.0087)	0.0033 (0.0082)	0.0027 (0.0089)
$\delta_1^{(2)}$	—	—	0.0156 (0.0062)	0.1025 (0.0182)	0.0988 (0.0226)
$\delta_2^{(2)}$	—	—	−0.0026 (0.0087)	−0.0122 (0.0242)	0.0174 (0.0311)
δ_1	—	0.9385 (0.1403)	—	—	0.4337 (0.0834)
δ_2	—	−0.9461 (0.2105)	—	—	−0.6798 (0.2732)
ν_1	3.4019 (1.4695)	3.0337 (1.1981)	3.4342 (1.8676)	5.7408 (5.5767)	2.4909 (0.3935)
ν_2	2.7620 (0.5728)	2.5497 (0.3183)	2.5194 (0.5290)	2.2577 (0.1821)	3.1796 (0.6434)
Q_{30}	27.0962 [0.4429]	39.8058 [0.1327]	30.2040 [0.5177]	33.1596 [0.3558]	41.2709 [0.1185]
Q_{30}^2	39.6623 [0.0753]	35.6250 [0.3810]	42.3384 [0.0986]	39.9753 [0.1279]	39.1482 [0.1320[
$\mathcal{L}_{\max}$	−191.184	−177.718	−186.350	−175.943	−165.534
AIC	412.369	389.437	[28]406.701	389.885	373.069
BIC	464.765	413.355	466.083	456.253	446.423

Figures in parentheses (square brackets) are standard errors (P-values).

- The Ljung–Box statistic based on standardized residuals (Q_r) with $r = 30$ lags; the Ljung–Box statistic based on squared standardized residuals (Q_r^2); the maximized log-likelihood ($\mathcal{L}_{\max}$); the Akaike information criterion (AIC); the Bayesian information criterion (BIC).
- Since the asymptotic distribution of residual autocorrelations from nonlinear models such as those considered here is not generally the same as that obtained under linearity (see, e.g., Li, 1992; Hwang et al., 1994), the P-values of Q_r and Q_r^2 are computed from a bootstrap approximation to their null sampling distributions instead of the usual chi-squared asymptotic approximation.

- Table 4 reports the ML estimation results for C-STAR(4), SDC-STAR(4), RC-STARX(4), C-STARX(4) and SDC-STARX(4) models (under the assumption of Student-t conditional distributions).
- For the SDC-STAR and the SDC-STARX models, the coefficients on output growth (δ_1) and inflation (δ_2) in the equation which determines the state-dependent threshold are both significantly different from zero. The results suggest that, for both models, the interest rate threshold values are such that **regime 1 is favoured the lower output growth is and the higher inflation is, while regime 2 is favoured for high values of output growth and low values of inflation.**
- This is because high values of inflation (and low values of output growth) result in small values for the threshold, which makes it more likely that the interest rate would exceed the threshold value. Figure 6, shows that the SDC-STAR gives weights close to unity in several periods which can be mostly associated with economic recessions. We find that the C-STAR is a not a valid restriction of the SDC-STAR. The AIC and BIC also favour the SDC-STAR model over the C-STAR.

State Dependent Threshold, Mixing Funtion for the SDC−STAR

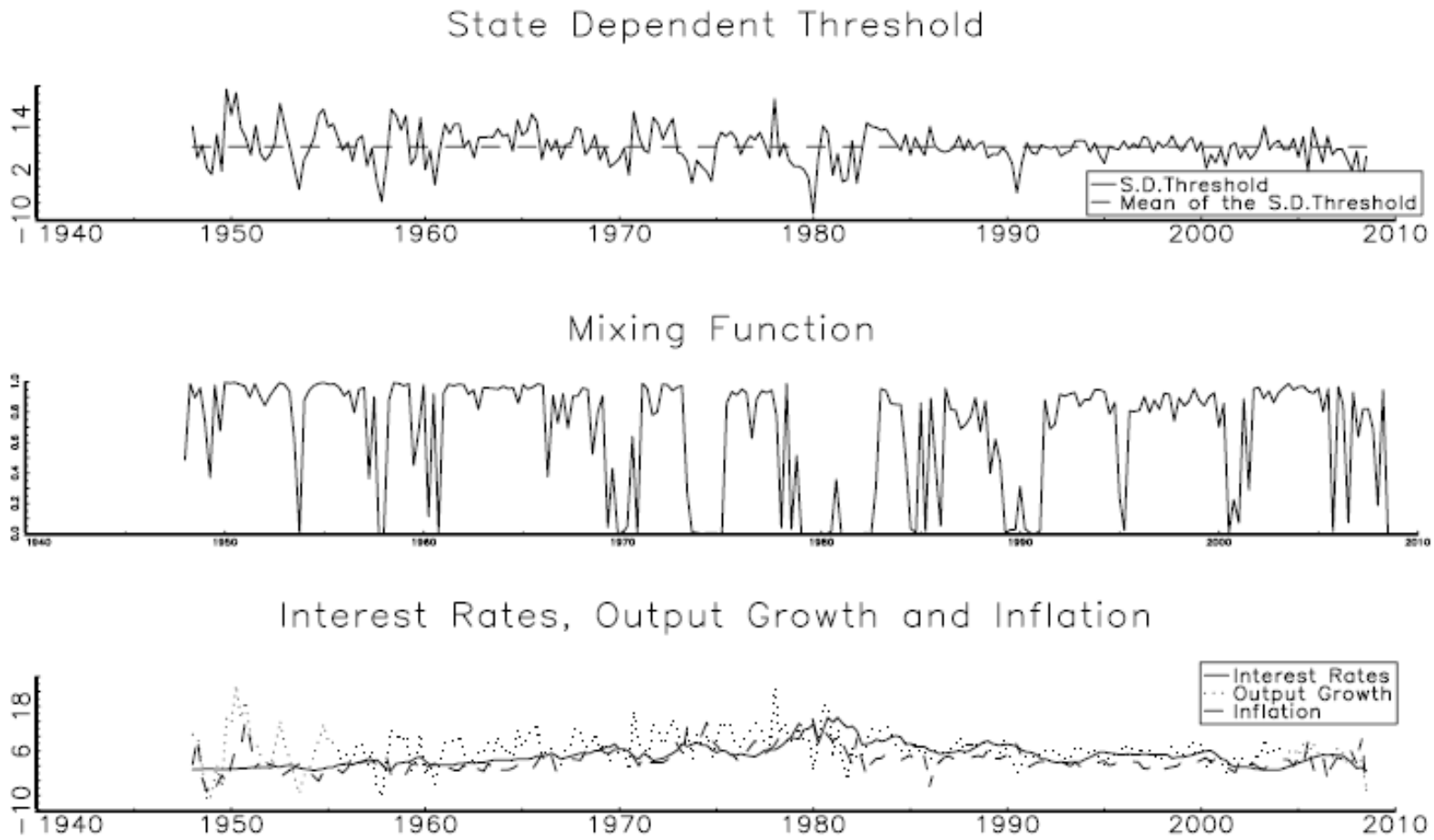


Figure 6

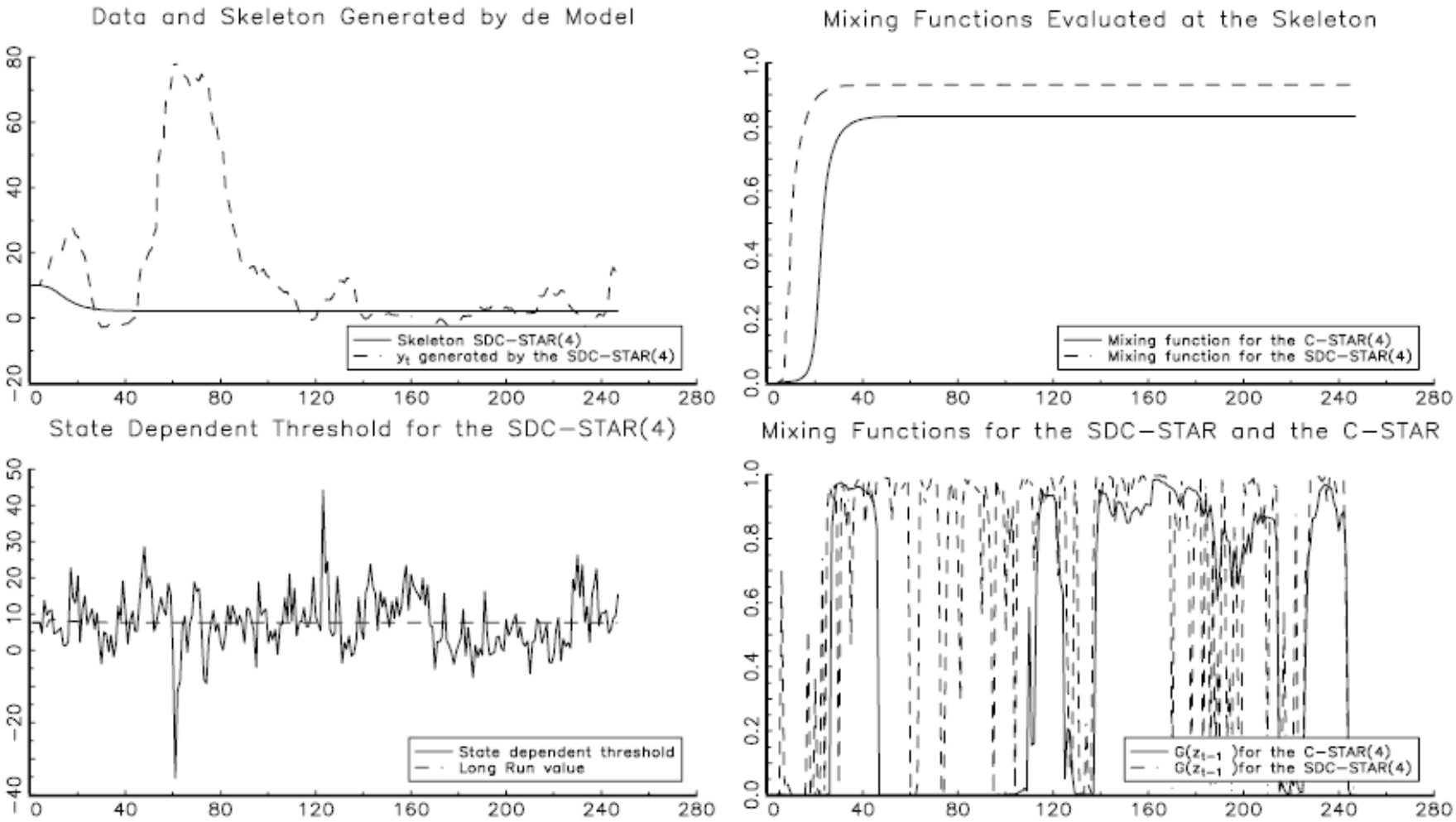

Data and Skeleton Generated by de Model
Skeleton SDC−STAR(4)
y_t generated by the SDC−STAR(4)
Mixing Functions Evaluated at the Skeleton
Mixing function for the C−STAR(4)
Mixing function for the SDC−STAR(4)
State Dependent Threshold for the SDC−STAR(4)
State dependent threshold
Long Run value
Mixing Functions for the SDC−STAR and the C−STAR
$G(z_{t-1})$ for the C−STAR(4)
$G(z_{t-1})$ for the SDC−STAR(4)

Figure 7

- The stability of the empirical SDC-STAR(4) is assessed by using numerical simulation. The skeleton of the model is found to have a single stable fixed point $Y_e = 2.2129$. The values of the skeleton are plotted in the top left panel of Figure 7, along with artificial data obtained by using the fitted model as the DGP.
- We note that the simulated data appear to replicate the qualitative features of the observed data – there is a fairly long period during which the series diverges from its long-run value, which takes the series above the threshold. The top right panel in Figure 7 shows the values of the mixing function $G(\cdot)$ evaluated under the skeleton of the SDC-STAR(4) and C-STAR(4) models, while the bottom left panel shows the evolution of the state-dependent threshold and its convergence to its long-run value 7.591. Finally, the bottom right panel shows the evolution of the mixing functions for the simulated data, which reveals substantial differences between the absolute and relative (to the exogenous variable) mixing functions.

- Within the class of SD-CSTARX models, there are several testable hypotheses of interest. We find that the C-STARX specification is not a valid reduction of the SD-CSTARX model since the LR statistic for $\delta_1 = \delta_2 = 0$ has a value of 20.818, leading to a rejection of the hypothesis of a constant threshold at the 1% significance level.
- The RSDC-STARX specification is not a valid simplification either since the LR statistic for $\delta_1^{(1)} - \delta_1^{(2)} = \delta_2^{(1)} - \delta_2^{(2)} = \delta = 0$ is 20.814. When the SD-CSTAR is tested against the SD-CSTARX, the LR statistic for $\delta_1^{(1)} = \delta_1^{(2)} = \delta_2^{(1)} = \delta_2^{(2)} = 0$ has a value of 24.368, implying that the model in which output growth and inflation are allowed to influence the regime-specific variables directly enjoys more support by the data.
- The SDC-STARX model is favoured by the AIC, while the SDC-STAR is favoured by the BIC. Overall, there is significant evidence in favour of specifications with a state-dependent threshold.
- The stability of the empirical SDC-STARX(4) is assessed in a similar way, with the corresponding plots shown in Figure 11. We find a single stable fixed point $Y_e = 5.50$ with a long-run value of 5.57 for the threshold.

State Dependent Threshold and Mixing Function for the SDC−STARX

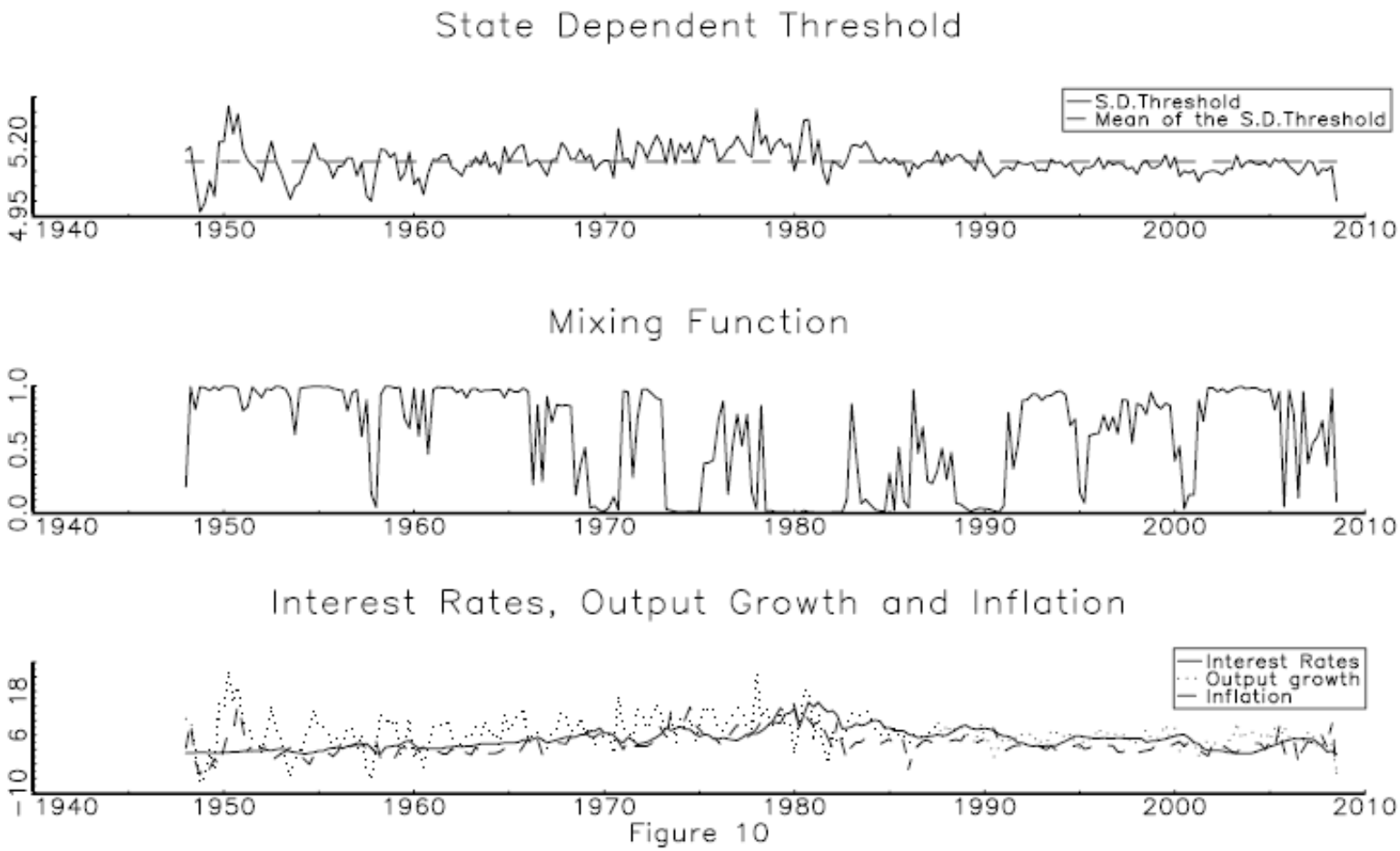


Figure 10

Generated Data and Skeleton for the Empirical DGP

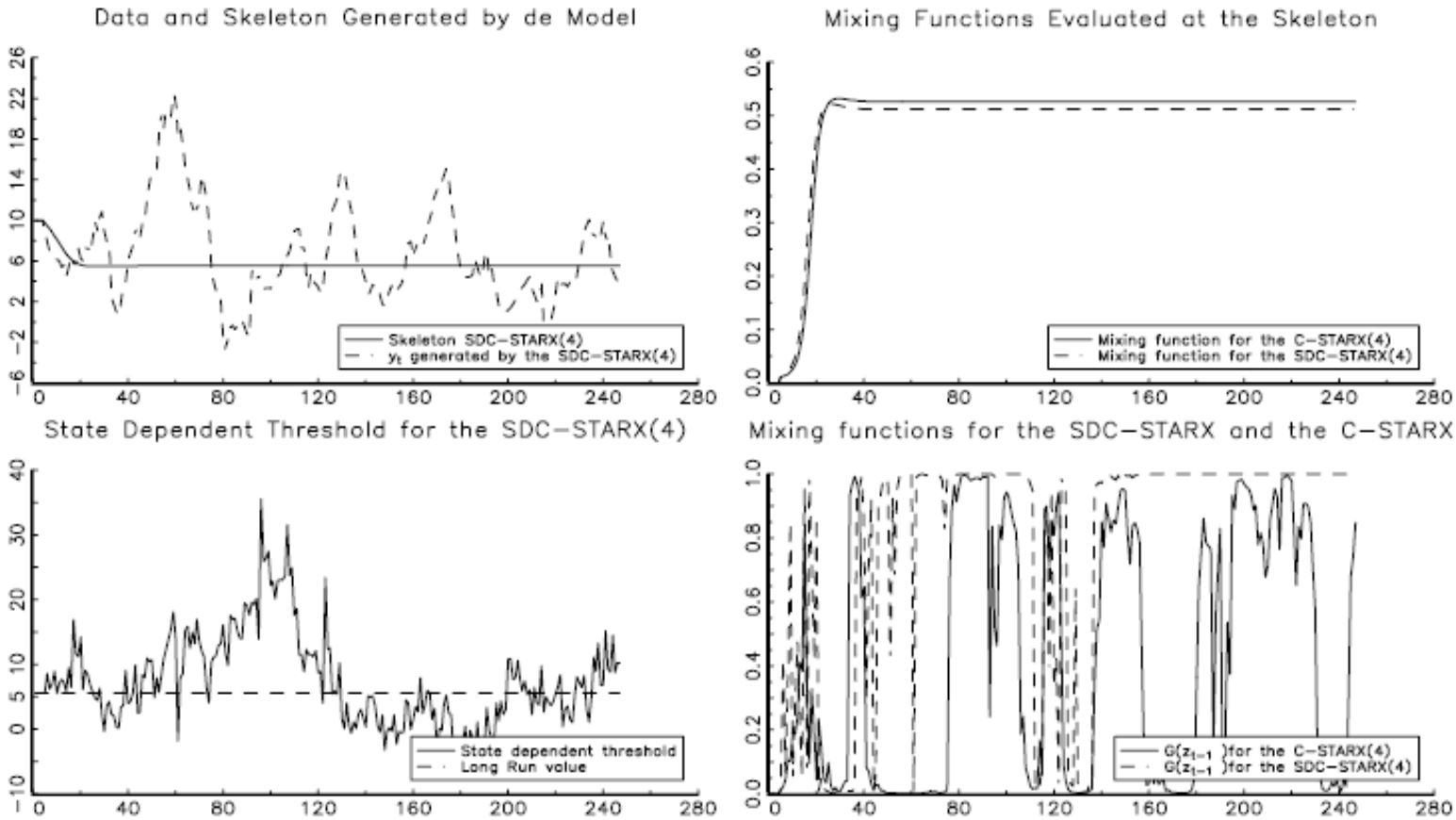


Figure 11

Summary

- In summary, the results presented in this section suggest that models with state-dependent thresholds, i.e., SDC-STAR and SDC-STARX models, are capable of characterizing the data successfully. In our application at least, what seems to matter is not so much the absolute value of the threshold but rather its level relative to other variables that dictate the evolution of the economy as a whole.
- We believe that, for economic variables such as those considered in this paper, a relative threshold with respect to other variables that dictate the evolution of the economy is of more importance for the dynamics of the variables under consideration than a fixed threshold. Furthermore, since constant-threshold specifications are special cases of state-dependent specifications, it seems reasonable to use state-dependent specifications as a first step in a modelling cycle.

Switching Regime Estimation

Series de TIempo

UTDT

January 2020

- The economy (the time series) often behaves very differently in periods such as booms and recessions. Markov Switching models are a useful way of characterizing this phenomena.

Uses of Markov Switching Models:

- Estimating models which are state dependent.
- Use the structure of the model to assess the probability that a state takes place, say a boom.
- Incorporate this feature in Rational Expectations models, pricing, derivatives, etc.

STRUCTURE OF THE LECTURE

- Definition and Properties.
- Estimation and testing
- Multivariate extensions and a R.E. example

Introduction

- Underlying assumption in all econometric models is that all observations have been drawn from the same distribution conditional on some constant parameter set
- The standard approach consists of trying to detect the existence of the regime changes and then imposing dummies.
  - Models with too many dummies

How should we model a change in the process?

- Suppose that the series under scrutiny has a break in its unconditional mean at time t_1. For data prior to t_1 we might use:

$$y_t - \mu_0 = \phi(y_{t-1} - \mu_0) + \varepsilon_t \qquad \text{for } t_1 < t$$

  and for data after t_1

$$y_t - \mu_1 = \phi(y_{t-1} - \mu_1) + \varepsilon_t \qquad \text{for } t \geq t_1$$

- Even if this captures the break, it is not a satisfactory model:
  - A complete time series model would include a description of the probability law governing the change from μ_0 to μ_1.

- We might consider the process to be influenced by an unobserved random variable x_t
  - x_t is the *state or regime*
- In the example above, we could regard x_t as:

$$x_t = \begin{cases} 0 \text{ if the process has mean } \mu_0 \\ 1 \text{ if the process has mean } \mu_1 \end{cases}$$

- Thus, we could write

$$y_t - \mu_{x_t} = \phi_1 (y_{t-1} - \mu_{x_{t-1}}) + \varepsilon_t$$

  where $\mu_{x_t} = (1 - x_t)\mu_0 + x_t \mu_1$
- Process for the unobserved variable: *Markov chain*

Properties of the Markov Process

Definition

Let x_t be a random variable that can take values 0 and 1. If the probability that x_t takes a particular value at time t, only depends on its value at $t-1$, this variable is governed by a *Markov process* of order 1.

$$P((x_t = i|x_{t-1} = j, x_{t-2} = k...) = P((x_t = i|x_{t-1} = j)$$

Thus the process is summarized by the probabilities:

$$P(x_t = 0|x_{t-1} = 0) = q \ , \ P(x_t = 1|x_{t-1} = 1) = p.$$

The transition Matrix

	0	1	(time t-1)
0	q	$(1-p)$	
1	$(1-q)$	p	
(time t)			

Autoregressive Representation of Markov Process

$$\begin{bmatrix} 1-x_{t+1} \\ x_{t+1} \end{bmatrix} = \begin{bmatrix} q & (1-p) \\ (1-q) & p \end{bmatrix} \begin{bmatrix} 1-x_t \\ x_t \end{bmatrix} + \begin{bmatrix} \zeta_{1,t+1} \\ \zeta_{2,t+1} \end{bmatrix}$$

- The second row gives

$$x_{t+1} = (1-q) + (-1+p+q)x_t + \zeta_{2,t+1}$$

- This expression can be recognized as an AR(1) process with constant term $1-q$ and autoregressive coefficient $(-1+p+q)$

- Expected value of x_t :

$$E(x_{t+1}) = (1-q) + (-1+p+q)E(x_t)$$

or

$$E(x_t) = \frac{1-q}{2-p-q}$$

since $E(x_{t+1}) = E(x_t)$ for a stationary process.

- **Unconditional probabilities** of being in state 1 and 0.
- Notice that $E(x_t) = 0P(x_t = 0) + 1P(x_t = 1) = \frac{1-q}{2-p-q}$.
- Then,

$$P(x_t = 1) = \frac{1-q}{2-p-q},$$

$$P(x_t = 0) = \frac{1-p}{2-p-q}.$$

Conditional and Unconditional Probabilities of States 0 and 1

An alternative derivation

- It can be shown that the unconditional probabilities at time zero multiplied by the matrix of transition probabilities are equal to the unconditional probabilities at time one:

$$\begin{bmatrix} P(x_1 = 0) \\ P(x_1 = 1) \end{bmatrix} = \begin{bmatrix} q & (1-p) \\ (1-q) & p \end{bmatrix} \begin{bmatrix} P(x_0 = 0) \\ P(x_0 = 1) \end{bmatrix}$$

Equilibrium Probabilities

- If the process is stationary, there exist state probabilities $\{\pi_0, \pi_1\}$ that satisfy:

$$\Pi = \mathbf{P}\Pi$$

where $\mathbf{P}$ is the matrix of transition probabilities.

$$\begin{bmatrix} \pi_0 \\ \pi_1 \end{bmatrix} = \begin{bmatrix} q & (1-p) \\ (1-q) & p \end{bmatrix} \begin{bmatrix} \pi_0 \\ \pi_1 \end{bmatrix}$$

where $\pi_0 = P(x_{t-j} = 0)$ and $\pi_1 = P(x_{t-j} = 1)$ for all values of j.

- Using $\pi_0 + \pi_1 = 1$:

$$\pi_0 = \frac{(1-p)}{(2-p-q)} \text{ and } \pi_1 = \frac{(1-q)}{(2-p-q)}$$

where π_0 and π_1 are the equilibrium unconditional probabilities.

- Given the following initial values:

$$\begin{bmatrix} p^0(0) \\ p^0(1) \end{bmatrix} = \begin{bmatrix} \pi_0 \\ \pi_1 \end{bmatrix}$$

  It can be shown (by multiplying n times by the transition probability matrix) that the unconditional probability vector at time n is:

$$\begin{bmatrix} p^n(0) \\ p^n(1) \end{bmatrix} = \begin{bmatrix} \pi_0 \\ \pi_1 \end{bmatrix}$$

- Therefore, the distribution does not change with time and the stochastic process is always in equilibrium.

Forecasts for a Markov Chain

- A n- period ahead forecast for a Markov chain can be obtained simply by multiplying n times by the transition probability:

$$\begin{bmatrix} P(x_{t+n}=0) \\ P(x_{t+n}=1) \end{bmatrix} = \begin{bmatrix} q & (1-p) \\ (1-q) & p \end{bmatrix}^n \begin{bmatrix} P(x_t=0) \\ P(x_t=1) \end{bmatrix}$$

$\mathbf{P}^n$ is derived in the following way:

1) Find the eigenvalues of the transition probability Matrix.

$$\lambda_1 = 1, \quad \lambda_2 = -1 + p + q,$$

2) Find the associated eigenvectors.

$$\begin{bmatrix} \frac{(1-p)}{(2-p-q)} \\ \frac{(1-q)}{(2-p-q)} \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

3) Express $\mathbf{P}$ as $T\Lambda T^{-1}$, where

$$T = \begin{bmatrix} \frac{(1-p)}{(2-p-q)} & -1 \\ \frac{(1-q)}{(2-p-q)} & 1 \end{bmatrix}$$

is the matrix of eigenvectors and

$$\Lambda = \begin{bmatrix} 1 & 0 \\ 0 & -1+p+q \end{bmatrix}$$

a diagonal matrix of eigenvalues.

4) Use the result that

$$\mathbf{P}^n = T\Lambda^n T^{-1}$$

or

$$P^n = \begin{bmatrix} \frac{1-p}{(2-p-q)} & -1 \\ \frac{1-q}{2-p-q} & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1+p+q \end{bmatrix}^n \begin{bmatrix} \frac{1-p}{(2-p-q)} & -1 \\ \frac{1-q}{2-p-q} & 1 \end{bmatrix}^{-1}$$

(t+n) \ (t)	0	1
0	$\frac{(1-p)}{(2-p-q)} + \frac{(1-q)(p+q-1)^n}{(2-p-q)}$	$\frac{(1-p)}{(2-p-q)} - \frac{(1-p)(p+q-1)^n}{(2-p-q)}$
1	$\frac{(1-q)}{(2-p-q)} - \frac{(1-q)(p+q-1)^n}{(2-p-q)}$	$\frac{(1-q)}{(2-p-q)} + \frac{(1-p)(p+q-1)^n}{(2-p-q)}$

Note that by making $n = 1$ in the above matrix we end with the matrix of transition probabilities, $\mathbf{P}$.

- In addition, we can derive the conditional expectations:

$$E(x_{t+n}|x_t = 0) = \frac{(1-q)}{(2-p-q)} - \frac{(1-q)(p+q-1)^n}{(2-p-q)}$$

$$E(x_{t+n}|x_t = 1) = \frac{(1-q)}{(2-p-q)} + \frac{(1-p)(p+q-1)^n}{(2-p-q)}$$

- The expected value of x_t at time n conditional on s at time zero is:

$$E(x_{t+n}|x_t = x_t) = \frac{(1-q)}{(2-p-q)} + (x_t - \frac{(1-q)}{(2-p-q)})(p+q-1)^n$$

- This result is derived assuming that x_t is observed (conditional on x_t)

A brief description of Hamilton's non linear filter

- The procedure assumes that discrete states of the economy are not known: inferred from the data.
- States follow a discrete Markov process.

A brief description of Hamilton's non linear filter

- The observed variable, y_t , is assumed to follow an $AR(m)$:

$$y_t - \mu_{x_t} = \phi_1(y_{t-1} - \mu_{x_{t-1}}) + .. + \phi_m(y_{t-m} - \mu_{x_{t-m}}) + \sigma_{x_t}\varepsilon_t$$

  ε_t is distributed N(0,1) and μ_{x_t} is parameterized as $\alpha_0 + \alpha_1 x_t$ and σ_{x_t} as $w_0 + w_1 x_t$
- The error ε_t is assumed to be independent of all $x_{t-j} \geq 0$

Hamilton's Filter

Step 1

Step 1 Calculate the joint density of the m past states and the current state conditional on the information included in y_{t-1} and all past values of y, where y is the variable that is observed:

$$p(x_t, x_{t-1}, , ..., x_{t-m}|y_{t-1}, .., y_0) = p(x_t|x_{t-1})p(x_{t-1}, , ..., x_{t-m}|y_{t-1}, ..., y_0)$$

Step 1

As in all the subsequent steps the second term on the right-hand-side is obtained from the preceding step of the filter. In this case,
$p(x_{t-1}, x_{t-2}, ..., x_{t-m}|y_{t-1}, y_{t-2}, ..., y_0)$ is known from the input to the filter, which in turn represents the result of the iteration at date $t-1$ (from step 5).

Step 1

To begin with the iteration, it is necessary to assign some initial values to the parameters, and to impose some initial conditions on the Markov process. The unconditional distribution $p(x_{m-1}, x_{m-2}, ...x_0)$ has been chosen for the first observation:

$$p(x_{m-1}, x_{m-2}, ...x_0) = p(x_{m-1}|x_{m-2})...p(x_1|x_0)p(x_0)$$

where $p(x_0)$ are the equilibrium unconditional probabilities as defined above.

Step 2

Calculate the joint conditional distribution of y_t and $(x_t, x_{t-1}, ..., x_{t-m})$.

$$
\begin{aligned}
& p(y_t, x_t, x_{t-1}, ..., x_{t-m} | y_{t-1}, y_{t-2}, ..., y_0) \\
= \quad & p(y_t | x_t, x_{t-1}, ..., x_{t-m}, y_{t-1}, y_{t-2}, ..., y_0) \times \\
& p(x_t, x_{t-1}, x_{t-2}, ..., x_{t-m} | y_{t-1}, y_{t-2}, ..., y_0)
\end{aligned}
$$

Hamilton's Filter

Step 2

We assume that

$$
\begin{aligned}
& p(y_t | x_t, x_{t-1}, .., x_{t-m}, y_{t-1}, y_{t-2}, .., y_0) \\
= \quad & \frac{1}{\sqrt{2\pi}(\omega_0 + \omega_1 x_t)} \exp[-\frac{1}{2[\omega_0 + \omega_1 x_t]^2}((y_t - \alpha_1 x_t - \alpha_0) \\
& -\phi_1(y_{t-1} - \alpha_1 x_{t-1} - \alpha_0) - .. - \phi_m(y_{t-m} - \alpha_1 x_{t-m} - \alpha_0))^2]
\end{aligned}
$$

Step 3

Marginalize the previous joint density with respect to the states giving the conditional density, from which the (conditional) likelihood function is calculated.

$$
\begin{aligned}
& p(y_t | y_{t-1}, y_{t-2}, .., y_0) \\
= & \sum_{x_t=0}^{1} \sum_{x_{t-1}=0}^{1} .. \sum_{x_{t-m}=0}^{1} p(y_t, x_t, x_{t-1}, .., x_{t-m} | y_{t-1}, y_{t-2}, .., y_0)
\end{aligned}
$$

Step 4

Combine the results from steps 2 and 3 to calculate the joint density of the state conditional on the observed current and past realizations of y

$$
\begin{aligned}
& p\left(x_t, x_{t-1}, ., x_{t-m} \middle| y_t, y_{t-1}, y_{t-2}, ., y_0\right) \\
= \quad & \frac{p\left(y_t, x_t, x_{t-1}, ., x_{t-m} \middle| y_{t-1}, y_{t-2}, ., y_0\right)}{p\left(y_t \middle| y_{t-1}, y_{t-2}, ., y_0\right)}
\end{aligned}
$$

Step 5

The desired output is then obtained from

$$p(x_t, .., x_{t-m+1}|y_t, y_{t-1}.., y_0) = \sum_{x_{t-m}=0}^{1} p(x_t, .., x_{t-m}|y_t, .., y_0)$$

The output of step 5 is used as an input to the filter in the next iteration.

Hamilton's Filter

Step 5

- Note that to iterate, estimates of the parameters are required.
- Maximum likelihood estimates can be obtained numerically from Step 3 as a by-product of the filter

$$\ln p(y_t, y_{t-1}, y_{t-2}, .., y_m | y_{m-1}, .., y_0) = \sum_{t=m}^{T} \ln p(y_t | y_{t-1}, .., y_0).$$

which can be maximized numerically with respect to the unknown parameters $(\alpha_1, \alpha_0, p, q, \omega_0, \omega_1, \phi_1, \phi_2 ... \phi_m)$.

Comments

- Notice that p and q, the parameters of the transition matrix, are also estimated by maximum likelihood.
- Hamilton's filter requires the numerical optimization of a very complicated non-linear function.

Specification Tests

- Specification tests based on the properties of the standardized residuals, $\widehat{\varepsilon}_t = \frac{y_t - \widehat{E(y_t|I_{t-1})}}{\widehat{\sigma}_t}$.
- First calculate the of y_t given information at time $t-1$:

$$\begin{aligned} E(y_t|I_{t-1}) &= \alpha_0 + \alpha_1 E(x_t|I_{t-1}) + \phi_1(y_{t-1} - \alpha_1 E(x_{t-1}|I_{t-1}) - \alpha_0) \\ &\quad + ... + \phi_m(y_{t-m} - \alpha_1 E(x_{t-m}|I_{t-1}) - \alpha_0), \end{aligned}$$

where

$$E(x_t|I_{t-1}) = \frac{(1-q)}{(2-p-q)} + (P(x_{t-1}|I_{t-1}) - \frac{(1-q)}{(2-p-q)})(p+q-1),$$

$$E(x_{t-m}|I_{t-1}) = P(x_{t-m} = 1|I_{t-1}), m \geq 1.$$

- $P(x_{t-m}|I_{t-1})$, for $m > 1$, are called "*smoothing probabilities*"
  - They can be calculated from the "filtering probabilities".

Computing the standard deviation

Step 1 First make use of the autoregressive representation of the Markov process

$$x_t = (1-q) + (-1+p+q)x_{t-1} + \zeta_{2,t}$$

For this process the error, conditional on $x_{t-1} = 1$, can be characterized as

$$\zeta_{2,t} = \begin{array}{l} (1-p) \text{ with probability } p \\ -p \text{ with probability } 1-p \end{array}$$

and conditional on $x_{t-1} = 0$

$$\zeta_{2,t} = \begin{array}{l} -(1-q) \text{ with probability } q \\ q \text{ with probability } 1-q \end{array}$$

Computing the standard deviation

Step 2 Calculate the variance of the error term, $\zeta_{2,t}$, conditional on the state at $t-1$.

$$
\begin{aligned}
E(\zeta_{2,t}^2 | x_{t-1} &= 1) = (1-p)^2 p + p^2 (1-p) = p(1-p) \\
E(\zeta_{2,t}^2 | x_{t-1} &= 0) = (1-q)^2 q + q^2 (1-q) = q(1-q)
\end{aligned}
$$

Computing the standard deviation

Step 3 Calculate the conditional variance (conditional on $I_{t-1} = \{y_{t-1}, \ldots y_0\}$)

We start by calculating the state dependent variance $\sigma^2_{x_t}$ as a function of the Markov switching parameters.

Computing the standard deviation

Step 3 Conditional on $x_{t-1} = 1$, the switching variance can be written as:

$$\sigma^2_{x_t} = E(\sigma^2_{x_t}|x_{t-1} = 1) + V(\mu_{x_t}|x_{t-1} = 1)$$

where

$$E(\sigma^2_{x_t}|x_{t-1} = 1) = (E(\sigma_{x_t}||x_{t-1} = 1))^2 + Var((\sigma_{x_t}|x_{t-1} = 1))$$

(since σ_{x_t} is a random variable) and

$$Var((\sigma_{x_t}|x_{t-1} = 1)) = E(\sigma^2_{x_t}|x_{t-1} = 1) - (E(\sigma_{x_t}|x_{t-1} = 1))^2$$

Computing the standard deviation

Step 3 Then using that

$$(E(\sigma_{x_t}|x_{t-1}=1))^2 = (w_0 + w_1 E(x_t|x_{t-1}=1)^2 = (w_0 + w_1 p)^2$$

$$\begin{aligned} Var((\sigma_{x_t}|x_{t-1} &= 1)) = Var(w_0 + w_1 x_t|x_{t-1}=1) \\ &= w_1^2 p(1-p). \end{aligned}$$

$$\begin{aligned} V(\mu_{x_t}|x_{t-1} &= 1) = Var(\alpha_0 + \alpha_1 x_t|x_{t-1}=1) \\ &= \alpha_1^2 p(1-p) \end{aligned}$$

Computing the standard deviation

Step 3 Collecting all these terms we can see that

$$\sigma^2_{x_t} = (w_0 + w_1 p)^2 + w_1^2 p(1-p) + \alpha_1^2 p(1-p)$$

We can obtain a similar formulae for the variance conditional on $x_{t-1} = 0$.

$$\sigma^2_{x_t} = E(\sigma^2_{x_t} | x_{t-1} = 0) + V(\mu_{x_t} | x_{t-1} = 0)$$

and doing the same transformations for state 0 we obtain

$$\sigma^2_{x_t} = (w_0 + w_1(1-q))^2 + w_1^2 q(1-q) + \alpha_1^2 q(1-q).$$

Computing the standard deviation

Step 3 Clearly the state is not observed at time $t-1$ but we can use the filtered probabilities to make an inference of the unobserved state.

Then the conditional variance (on information on time $t-1$) is

$$\sigma_t^2 = \begin{array}{c} ((w_0 + w_1 p)^2 + w_1^2 p(1-p) \\ +\alpha_1^2 p(1-p)) P(x_{t-1} = 1 | I_{t-1}) + \\ ((w_0 + w_1(1-q))^2 + w_1^2 q(1-q) + \\ \alpha_1^2 q(1-q))(1 - P(x_{t-1} = 1 | I_{t-1})) \end{array} .$$

Computing the standard deviation

Step 3 Then the standardized residuals are simply, $v_t = \varepsilon_t / \sigma_t$ and we may conduct standard specification tests for these residuals.

Number of tests and specification tests

- Crucial: rightly identify the number of states or regimes.
- Hamilton proposes to use simple specification tests as a mean of assessing whether the estimated equation contains the right number of states.

A Bivariate model with regime switching

- We consider a VAR process in two variables, with 1 lag, with the feature that the means of each equation and the variance-covariance matrix are allowed to switch endogenously between two possible states.
- The two equations that define the VAR are influenced by the same state variable.
- The state is not observed and has to be inferred from a filter.

$$S_t' = \Phi S_{t-1}' + \psi D_{t-1}' + (\omega_0 + \omega_1 x_t)\nu_t$$

$$D_t' = \varphi S_{t-1}' + \Omega D_{t-1}' + (\tau_0 + \tau_1 x_t)\varepsilon_t$$

A Bivariate model with regime switching

- The centered variables are defined by the two following equations:

$$S'_t = S_t - \alpha_0 - \alpha_1 x_t$$

$$D'_t = D_t - \beta_0 - \beta_1 x_t$$

- A prime (') is used to denote centred variables in the remainder of the presentation
- x_t denotes the unobserved state of the system and takes values 0 and 1.

A Bivariate model with regime switching

- x_t is governed by a Markov process

$$\begin{aligned} P(x_t &= 0 | x_{t-1} = 0) = q \\ P(x_t &= 1 | x_{t-1} = 1) = p \end{aligned}$$

- The errors ν_t, ε_t are assumed to be independent of all x_{t-j}. $j \geq 0$.

A Bivariate model with regime switching

Substituting the centered variables into the VAR and rearranging terms, we obtain the following expression for S_t and D_t

$$\begin{aligned} S_t &= \alpha_0(1-\Phi) - \beta_0\psi + \Phi S_{t-1} + \psi D_{t-1} \\ &\quad + \alpha_1(x_t - \Phi x_{t-1}) - \beta_1\psi x_{t-1} + (\omega_0 + \omega_1 x_t)\nu_t \end{aligned}$$

$$\begin{aligned} D_t &= -\alpha_0\varphi + \beta_0(1-\Omega) + \varphi S_{t-1} + \Omega D_{t-1} \\ &\quad - \alpha_1\varphi x_{t-1} + \beta_1(x_t - \Omega x_{t-1}) + (\tau_0 + \tau_1 x_t)\varepsilon_t \end{aligned}$$

Testing The Term Structure of Interest Rates

- The process that drives the spread and the short-term interest rate difference is the VAR of equation described above in which D_t denotes the first difference of the three month rate, $R_{1t} - R_{1t-1}$ and S_t denotes the yield spread $R_{2t} - R_{1t}$.
- The expectations hypothesis of the term structure of the interest rates can be written as

$$S_t = (1/2)E_t D_{t+1} + \theta + u_t$$

- The restrictions imposed by the expectations model are presented below. Both an unrestricted and a restricted VAR can be estimated, and the restrictions tested using a likelihood ratio test.

Derivation of the restrictions in the regime-shifting VAR.

$$\begin{bmatrix} S_t' \\ D_t' \end{bmatrix} = \begin{bmatrix} \Phi & \psi \\ \varphi & \Omega \end{bmatrix} \begin{bmatrix} S_{t-1}' \\ D_{t-1}' \end{bmatrix} + \begin{bmatrix} (\omega_0 + \omega_1 x_t)\nu_t \\ (\tau_0 + \tau_1 x_t)\varepsilon_t \end{bmatrix}$$

To find the restrictions we condition on information available at $t-1$ on both sides of the term structure equation.

$$E(S_t | I_{t-1}^*) = (1/2)E(D_{t+1} | I_{t-1}^*) + \theta$$

were $I_{t-1}^* = \{S_{t-1}, S_{t-2}, .., D_{t-1}, D_{t-2}, .., x_{t-1}, .., \}$.

Then, we need to calculate expected values $E[D_{t+1}|I^*_{t-1}]$ and $E[S_t|I^*_{t-1}]$. These can be calculated in the following way:

$$E[D_{t+1}|I^*_{t-1}] = \beta_0 + \beta_1 E(x_{t+1}|I^*_{t-1}) + E(D'_{t+1}|I^*_{t-1}),$$

where

$$\begin{aligned} E[D'_{t+1}|I^*_{t-1}] &= [0,1]\,\Delta^2 Z'_{t-1}, \\ E(x_{t+1}|I^*_{t-1}) &= [\rho + (x_{t-1} - \rho)\lambda^2] \end{aligned}$$

where $\rho = \frac{1-q}{2-p-q}$ and $\lambda = (p+q-1)$.

$$E[S_t|I^*_{t-1}] = \alpha_0 + \alpha_1 E(x_t|I^*_{t-1}) + E(S'_t|I^*_{t-1})$$

where

$$\begin{aligned} E(S'_t|I^*_{t-1}) &= \begin{bmatrix} 1 & 0 \end{bmatrix} \Delta Z'_{t-1} \\ E(x_t|I^*_{t-1}) &= [\rho + (x_{t-1} - \rho)\lambda] \end{aligned}$$

Finally, substituting the expected (difference of the) short-term interest rates and the expected spread, the term structure of interest rates relationship can be expressed as;

$$\begin{aligned} & \alpha_0 + \alpha_1 [\rho + (x_{t-1} - \rho)\lambda] + [1, 0]\, \Delta Z'_{t-1} \qquad ((A4)) \\ = \quad & \frac{1}{2}[\beta_0 + \beta_1 [\rho + (x_{t-1} - \rho)\lambda^2] + [0, 1]\, \Delta^2 Z'_{t-1}] + \theta \end{aligned}$$

The restrictions written out in full are:

$$\begin{aligned} \alpha_1 &= \frac{1}{2}\beta_1 \lambda \\ \Phi &= \frac{\varphi \Omega}{2 - \varphi} \\ \psi &= \frac{\Omega^2}{2 - \varphi} \end{aligned}$$

Contagion

- Define:

$$z_t = [x_t, y_t]'$$

$$z_t = \mu + \Phi_{s_t} u_t$$

  where $\mu = [\mu_x, \mu_y]'$ and u_t is a Gaussian process

- $\{s_t\}$ is modelled as a time-homogeneous Markov chain on $\{1, 2, 3, 4\}$, independent of $\{u_t\}$
- Thus,

$$z_t | (s_t = s) \sim N(\mu, \Omega_{s_t})$$

Contagion

- The variance covariance matrix are:

$$\Omega = \left\{ \Omega_{s=1} = \begin{bmatrix} \sigma^2_{xh} & \sigma_{xh,yh} \\ \sigma_{yh,xh} & \sigma^2_{yh} \end{bmatrix}, \; \Omega_{s=2} = \begin{bmatrix} \sigma^2_{xh} & \sigma_{xh,yl} \\ \sigma_{yl,xh} & \sigma^2_{yl} \end{bmatrix} \right.$$
$$\left. \Omega_{s=3} = \begin{bmatrix} \sigma^2_{xl} & \sigma_{xl,yh} \\ \sigma_{yh,xl} & \sigma^2_{yh} \end{bmatrix}, \; \Omega_{s=4} = \begin{bmatrix} \sigma^2_{xl} & \sigma_{xl,yl} \\ \sigma_{yl,xl} & \sigma^2_{yl} \end{bmatrix} \right\}$$

Contagion

- Transition matrix: 4×4 matrix Π (with elements $\pi_{ij} = \Pr(s_t = i | s_{t-1} = j),\ i, j = 1, 2, 3, 4$)
- No contagion amounts to:

$$\begin{pmatrix} \pi_{xh}\pi_{yh} & \pi_{xh}(1-\pi_{yl}) & (1-\pi_{xl})\pi_{yh} & (1-\pi_{xl})(1-\pi_{yl}) \\ \pi_{xh}(1-\pi_{yh}) & \pi_{xh}\pi_{yl} & (1-\pi_{xl})(1-\pi_{yh}) & (1-\pi_{xl})\pi_{yl} \\ (1-\pi_{xh})\pi_{yh} & (1-\pi_{xh})(1-\pi_{yl}) & \pi_{xl}\pi_{yh} & \pi_{xl}(1-\pi_{yl}) \\ (1-\pi_{xh})(1-\pi_{yl}) & (1-\pi_{xh})\pi_{yl} & \pi_{xl}(1-\pi_{yh}) & \pi_{xl}\pi_{yl} \end{pmatrix}.$$

Contagion

- Contagion will occur when one of the countries leads (or lags) the other one.
- This hypothesis can be verified testing (using LR tests distributed as $\chi^2(10)$) if we can reduce the transition matrices to $\Pi_1^{xly} =$

$$\begin{pmatrix} \pi_{xh} & \pi_{xh} & 0 & 0 \\ 0 & 0 & (1-\pi_{xl}) & (1-\pi_{xl}) \\ (1-\pi_{xh}) & (1-\pi_{xh}) & 0 & 0 \\ 0 & 0 & \pi_{xl} & \pi_{xl} \end{pmatrix}$$

where Π_1^{xly} indicates x leads y one period

Markov Switching Causality

- The analysis of Granger causality between x_1 and x_2 is based on the following Markov switching VAR model:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} \mu_{10} + \mu_{11} s_{1,t} \\ \mu_{20} + \mu_{21} s_{2,t} \end{bmatrix} + \begin{bmatrix} \phi_{10} + \phi_{11} s_{1,t} & \psi_1 s_{1,t} \\ \psi_2 s_{2,t} & \phi_{20} + \phi_{21} s_{2,t} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix}$$
$$+ \begin{bmatrix} \varphi_{10} + \varphi_{11} s_{1,t} \\ \varphi_{20} + \varphi_{21} s_{2,t} \end{bmatrix} z_{t-1} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}, \quad t = 1, \ldots, T.$$

  where $s_{1,t}$ $s_{2,t} \epsilon \{0, 1\}$ are unobserved random variables
- x_2 Granger causes x_1 when $s_{1,t} = 1$ and is Granger non-causal for x_1 when $s_{1,t} = 0$

Target Zone Credibility

- Dynamic of the nominal interest rate differential

$$r_t - \mu(s_t) = \sum_{j=1}^{m} \phi_j [r_{t-j} - \mu(s_{t-j})] + \sigma(s_t)\varepsilon_t, \qquad (t = 1, \ldots, T)$$

where $s_t \epsilon \{0, 1\}$ is the state variable with

$$\begin{aligned}
&\Pr(s_t = 1 | s_{t-1} = 1) = p, \\
&\Pr(s_t = 0 | s_{t-1} = 1) = 1 - p, \\
&\Pr(s_t = 0 | s_{t-1} = 0) = q, \\
&\Pr(s_t = 1 | s_{t-1} = 0) = 1 - q,
\end{aligned}$$

Target Zone Credibility

- $$\mu(s_t) = \alpha_0 + \alpha_1 s_t,$$

  and

  $$\sigma(s_t) = \sigma_0(1 - s_t) + \sigma_1 s_t.$$

  where $\alpha_1 - \alpha_0 > 0$

Target Zone Credibility

- To assess the links between target-zone credibility and macroeconomic variables:

$$\Pr(s_t = 1|s_{t-1} = 1, z_{t-1}) = p_t = \exp(c_1 + \beta_1 z_{t-1})/[1 + \exp(c_1 + \beta_1 z_{t-1})].$$
$$\Pr(s_t = 0|s_{t-1} = 0, z_{t-1}) = q_t = \exp(c_0 + \beta_0 z_{t-1})/[1 + \exp(c_0 + \beta_0 z_{t-1})].$$
$$\Pr(s_t = 0|s_{t-1} = 1, z_{t-1}) = 1 - \exp(c_1 + \beta_1 z_{t-1})/[1 + \exp(c_1 + \beta_1 z_{t-1})],$$
$$\Pr(s_t = 1|s_{t-1} = 0, z_{t-1}) = 1 - \exp(c_0 + \beta_0 z_{t-1})/[1 + \exp(c_0 + \beta_0 z_{t-1})],$$

where z_t is an economic variable that affects the state transition probabilities.

Target Zone Credibility

- Sample log-likelihood function

$$\log \mathfrak{L} = \sum_{t=m+1}^{T} \log f\left(r_t | \mathcal{F}_{t-1}\right)$$

where $\mathcal{F}_i = (r_1, z_1, \ldots, r_i, z_i)$ $(i \geqslant 1)$ and $f\left(r_t | \mathcal{F}_{t-1}\right)$ represents the conditional density of r_t given the set of information that is available at date $t-1$

Target Zone Credibility

- Inferences about the unobserved regimes $\{s_t\}$ may be made on the basis of the filter probabilities $\Pr(s_t|\mathcal{F}_t)$, obtained as:

$$\Pr(s_t|\mathcal{F}_t) = \sum_{s_{t-1}=0}^{1} \sum_{s_{t-2}=0}^{1} \cdots \sum_{s_{t-m}=0}^{1} \Pr(s_t, s_{t-1}, \ldots, s_{t-m}|\mathcal{F}_t).$$

Intrinsic Bubbles and Regime Switching

- $$P_t = e^{-r} E_t(D_t + P_{t+1}).$$
- Any rational bubble B_t in the stock price satisfies
$$B_t = e^{-r} E_t(B_{t+1}).$$
- The process that drives the log of dividends is assumed to be a random walk with drift μ :
$$d_{t+1} = \mu + d_t + \xi_{t+1},$$
where $\xi_{t+1} \backsim N(0, \sigma^2)$.
- The "intrinsic bubble":
$$B(D_t) = cD_t^{\lambda}.$$

Intrinsic Bubbles and Regime Switching

- The parameter λ is the positive root of the quadratic equation

$$\frac{\sigma^2}{2}\lambda^2 + \mu\lambda - r = 0$$

- The present value (denoted by P_t^{pv}) is proportional to dividends:

$$P_t^{pv} = kD_t,$$

where $k = \left(e^r - e^{(\mu+\frac{1}{2}\sigma^2)}\right)^{-1}$.

$$P_t = kD_t + cD_t^{\lambda}.$$

Intrinsic Bubbles and Regime Switching

- Evolution of real dividends

$$d_{t+1} = d_t + \mu_0(1 - s_{t+1}) + \mu_1 s_{t+1} + (\sigma_0(1 - s_{t+1}) + \sigma_1 s_{t+1})\varepsilon_{t+1}$$

  where s_{t+1} follow an homogenous first order Markov Process and ε_{t+1} is an *iid* variable
- $p(s_t = 1|s_{t-1} = 1) = p$ and $p(s_t = 0|s_{t-1} = 0) = q$.
- The fundamental value of the stock is

$$P_t = \begin{cases} k_0 D_t \text{ if } s_t = 0 \\ k_1 D_t \text{ if } s_t = 1 \end{cases}$$

- Then k_0 and k_1 satisfy

$$\begin{aligned} k_0 &= e^{-r}(1 + qk_0 a_0 + (1-q)k_1 a_1). \\ &\text{and} \\ k_1 &= e^{-r}(1 + pk_1 a_1 + (1-p)k_0 a_0) \end{aligned}$$

where $a_0 = e^{(\mu_0 + \frac{1}{2}\sigma_0^2)}$ and $a_1 = e^{(\mu_1 + \frac{1}{2}\sigma_1^2)}$.

Intrinsic Bubbles and Regime Switching

- Intrinsic Bubble:
$$B_t = c_i D_t^{\lambda} \text{ when } s_t = i$$
- It satisfies
$$B_t = e^{-r} E(B_{t+1} | \Omega_t)$$
- Putting everything together
$$P_{s_t} = P_{s_t}^{pv} + B_{s_t}(D_t)$$

where
$$P_{s_t}^{pv} = (k_0(1 - s_t) + k_1 s_t))D_t,$$

and
$$B_{s_t}(D_t) = (c_0(1 - s_t) + c_1 s_t))D_t^{\lambda}.$$

Instrumental Variables

- Standard CAPM

$$u'(C_t) = \beta E_t[(1 + r_{t+1})u'(C_{t+1})]$$

$$E_t\left[\left(\frac{F_t - S_{t+1}}{P_{t+1}}\right)\frac{u'(C_{t+1})}{u'(C_t)}\right] = 0$$

- When all the variables are jointly lognormally distributed, the equation may be rewritten as

$$f_t = E_t[s_{t+1}] + \tfrac{1}{2}\mathrm{Var}_t[s_{t+1}] + \mathrm{Cov}_t[R_{t+1}, s_{t+1}],$$

Instrumental Variables

- UFER Hypothesis

$$f_t = E_t[s_{t+1}]$$

and

$$s_{t+1} = E_t[s_{t+1}] + \eta_{t+1}$$

- These equations are often expressed as

$$\Delta s_{t+1} = \alpha + \beta(f_t - s_t) + e_{t+1}$$

Instrumental Variables

- Consumption

$$C_t = \mu_{x_t} + \sum_{j=1}^{h} \varphi_{j,x_t} C_{t-j} + \sigma_{x_t} \zeta_t,$$

where $\{\zeta_t\}$ is a white noise and $\{x_t\}$ are regime-indicator variables independent of $\{\zeta_t\}$

$$q = \Pr[x_t = 0 | x_{t-1} = 0], \qquad p = \Pr[x_t = 1 | x_{t-1} = 1]$$

Instrumental Variables

- Conditional on $x_t = 0$, the solution for the forward rate is

$$f_t = \begin{array}{c} E_t[s_{t+1}] + \frac{1}{2}\mathrm{Var}_t[s_{t+1}] + q\mathrm{Cov}_t[R^{(0)}_{t+1}, s_{t+1}] + \\ (1-q)\mathrm{Cov}_t[R^{(1)}_{t+1}, s_{t+1}] \end{array}$$

- Conditional on $x_t = 1$, the solution is

$$f_t = \begin{array}{c} E_t[s_{t+1}] + \frac{1}{2}\mathrm{Var}_t[s_{t+1}] + (1-p)\mathrm{Cov}_t[R^{(0)}_{t+1}, s_{t+1}] \\ +p\mathrm{Cov}_t[R^{(1)}_{t+1}, s_{t+1}] \end{array}$$

where $R^{(i)}_t = \ln(u'(C^{(i)}_{t+1})/P_{t+1})$ and
$C^{(i)}_{t+1} = \mu_i + \sum_{j=1}^{h} \varphi_{j,i} C_{t+1-j} + \sigma_i \zeta_{t+1}$ for $i \in \{0, 1\}$.

Instrumental Variables

- These equations yield:

$$f_t = E_t[s_{t+1}] + A_t^{(0)}(1 - x_t) + A_t^{(1)} x_t,$$

with $A_t^{(i)} = (1/2)\mathrm{Var}_t[s_{t+1}] + \mathrm{Cov}_t[R_{t+1}^{(i)}, s_{t+1}]$ for $i \in \{0, 1\}$

Introduction to Bayesian Econometrics

May 2020

Introduction to Bayesian Econometrics

- Consider the following regression Model

$$Y = X\beta + \varepsilon, \qquad \varepsilon \tilde{} N(0, \sigma^2 I)$$

Within the Bayesian framework the parameters $\theta = \{\beta, \sigma^2\}$ are treated as random variables. This parameters have probability distributions which reflect the knowledge of the researcher, before observing the sample on Y and X, about the parameters of the model.

- This probability distribution, denoted as $g(\theta)$ is called a prior distribution.

Once Y is observed, the researcher revises the distribution of the parameters by combining the prior distribution with the information obtained in the sample using Bayes Theorem.

We will define the following concepts:

- $f(Y|\theta)$ denotes the distribution of Y (from where we draw the data) given the parameters
- $h(\theta, Y)$ denotes the joint distribution of Y and θ.
- $f(Y)$ denotes the marginal distribution of Y
- $p(\theta|Y)$ denotes the posterior distribution of θ given Y

- Then we can write the joint distribution as

$$h(\theta, Y) = f(Y|\theta)g(\theta) = p(\theta|Y)f(Y)$$

which allows to write the posterior as

$$p(\theta|Y) = \frac{f(Y|\theta)g(\theta)}{f(Y)}$$

or equivalently

$$p(\theta|Y) \propto f(Y|\theta)g(\theta)$$

- Using the functional equivalence between $f(Y|\theta)$ and the likelihood $L(\theta|Y)$ we can express the posterior as

$$p(\theta|Y) \propto L(\theta|Y)g(\theta).$$

As we will see later on, the classical and the Bayesian approach are the same when the prior information is not available, that is, when the prior is diffuse or flat.

Which prior distributions should be used?

- There are groups of densities that may be easier to combine with the information of the likelihood. The natural conjugate priors are priors that once they are combined with the likelihood, they produce a posterior which has the same distribution as the prior.
- Example: Distribution of β assuming that σ^2 is known

Assume $\beta|\sigma^2 \tilde{} N(\beta_0, \Sigma_0)$ (a multivariate normal distribution) where β_0 and Σ_0 are known. Then the distribution can be written as

$$\begin{aligned} g(\beta|\sigma^2) &= (2\pi)^{-\frac{K}{2}} |\Sigma_0|^{.5} exp\left\{-\frac{1}{2}(\beta-\beta_0)'\Sigma_0^{-1}(\beta-\beta_0)\right\} \\ &\propto exp\left\{-\frac{1}{2}(\beta-\beta_0)'\Sigma_0^{-1}(\beta-\beta_0)\right\} \end{aligned}$$

where $(2\pi)^{-\frac{K}{2}} |\Sigma_0|^{.5}$ is a known constant.

The log likelihood

$$\begin{aligned} L(\beta|\sigma^2, Y) &= (2\pi\sigma^2)^{-\frac{T}{2}} exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\} \\ &\propto exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\}. \end{aligned}$$

where $(2\pi\sigma^2)^{-\frac{T}{2}}$ is a known constant.
Then the posterior is

$$\begin{aligned} p(\beta|\sigma^2, Y) &\propto g(\beta|\sigma^2)L(\beta|\sigma^2, Y) \\ &\propto exp\left\{ \begin{array}{l} -\frac{1}{2}(\beta - \beta_0)'\Sigma_0^{-1}(\beta - \beta_0) \\ -\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta) \end{array} \right\}. \end{aligned}$$

Rearranging terms it can be shown that the posterior is also normal, therefore the normal density is the natural conjugate prior for β.

- Posterior distribution of β

It can be shown that $\beta|\sigma^2, Y \tilde{} N(\beta_1, \Sigma_1)$ where

$$\begin{aligned} \beta_1 &= (\Sigma_0^{-1} + \sigma^{-2}X'X)^{-1}(\Sigma_0^{-1}\beta_0 + \sigma^{-2}X'Y) \\ &= (\Sigma_0^{-1} + \sigma^{-2}X'X)^{-1}(\Sigma_0^{-1}\beta_0 + \sigma^{-2}X'X\widehat{\beta}) \\ \Sigma_1 &= (\Sigma_0^{-1} + \sigma^{-2}X'X)^{-1} \end{aligned}$$

From the previous equation we can see that the posterior mean of β is an average of β_0 and $\widehat{\beta}$.

- Example: Distribution of σ^2 assuming that β is known

The natural conjugate prior for σ^2 is the inverted Gamma distribution (the natural conjugate prior for $\frac{1}{\sigma^2}$ is the Gamma distribution)[1].
Prior of $\frac{1}{\sigma^2}|\beta \tilde{} \Gamma(\frac{v_0}{2}, \frac{\delta_0}{2})$ where v_0 and δ_0 are known.
Then

$$\begin{aligned} g(\frac{1}{\sigma^2}|\beta) &\propto (\frac{1}{\sigma^2})^{\frac{v_0}{2}-1} \exp(-\frac{\delta_0}{2\sigma^2}) \\ & \textit{and} \\ L(\frac{1}{\sigma^2}|\beta, Y) &= (2\pi\sigma^2)^{-\frac{T}{2}} exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\} \\ &\propto (\sigma^2)^{-\frac{T}{2}} exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\}. \end{aligned}$$

$$g(W) \propto W^{\frac{v}{2}-1} \exp(-\frac{W\delta}{2})$$

with $E(W) = \frac{v}{\delta}$ and $V(W) = 2\frac{v}{\delta^2}$.

- The posterior, $\frac{1}{\sigma^2}|\beta, Y \tilde{} \Gamma(\frac{v_1}{2}, \frac{\delta 1}{2})$, is therefore

$$
\begin{aligned}
p(\frac{1}{\sigma^2}|\beta, Y) &\propto g(\frac{1}{\sigma^2}|\beta)\ L(\frac{1}{\sigma^2}|\beta, Y) \\
&\propto (\frac{1}{\sigma^2})^{\frac{v_0}{2}-1} \exp(-\frac{\delta_0}{2\sigma^2})(\sigma^2)^{-\frac{T}{2}} exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\} \\
&= (\frac{1}{\sigma^2})^{\frac{v_0}{2}+\frac{T}{2}-1} exp\left\{-\frac{1}{2\sigma^2}(\delta_0 + (Y - X\beta)'(Y - X\beta))\right\} \\
&= (\frac{1}{\sigma^2})^{\frac{v_1}{2}-1} exp\left\{-\frac{\delta_1}{2\sigma^2}\right\} \\
\textit{where } \delta_1 &= \delta_0 + (Y - X\beta)'(Y - X\beta) \text{ and } v_1 = v_0 + T.
\end{aligned}
$$

GIbbs- Sampling in Econometrics

- Gibbs-sampling is a Markov chain Monte-Carlo method for approximating the joint and marginal distributions by sampling from conditional distributions.
- Consider the following joint density

$$f(z_1, z_2, ..., z_k)$$

and that we are interested in obtaining characteristics of the marginal density

$$f(z_t) = \int ..... \int f(z_1, z_2, ..., z_k) dz_1 dz_2, ..dz_{t-1} dz_{t+1} ....., dz_k$$

such as the mean or the variance. This exercise may be, when possible, very difficult to perform

- Gibbs sampling will allow me, if we are given the complete set of conditional distributions $f(z_t | z_1, z_2, ..z_{t-1}, z_{t+1} ....., z_k)$, to generate a sample of $z_1, z_2, ..., z_k$ without the need of knowing the joint $f(z_1, z_2, ..., z_k)$ or the marginals $f(z_t)$.

Methodology

Given arbitrary starting values $z_2^0, ..z_t^0, z_{t+1}^0 ....., z_k^0$

1) Draw z_1^1 from $f(z_1|z_2^0, ..z_t^0, z_{t+1}^0 ....., z_k^0)$
2) Then draw z_2^1 from $f(z_2|z_1^1, z_3^0, ..z_t^0, z_{t+1}^0 ....., z_k^0)$
3) Then draw z_3^1 from $f(z_3|z_1^1, z_2^1, z_4^0, z_5^0 ....., z_k^0)$

.

.

.

k) Finally draw z_k^1 from $f(z_k|z_1^1, z_2^1, z_3^1, z_4^1, z_5^1 ....., z_{k-1}^1)$

Then steps 1 to k can be iterated J times to get $z_1^j, z_2^j, ..z_t^j, z_{t+1}^j ....., z_k^j$, *for* $j = 1, 2, ..., J$.

- A crucial result in the literature is that of Geman and Geman (1984) that shows that the joint and marginal distributions of $z_1^j, z_2^j, ..z_t^j, z_{t+1}^j ....., z_k^j$ converge to the joint and marginal distributions of $z_1, z_2, ..z_t, z_{t+1} ....., z_k$ as $J \longrightarrow \infty$.
- Consider $J = L + M$, then typically what is done is to use the first L simulations until the Gibbs sampler has converged and then use the remaining M simulations to approximate the empirical distribution.
- Convergence of the Gibbs Sampling

The Convergence of the Gibbs sampler is a very important issue which is somehow difficult to handle. For example it is usual to plot the posterior densities over the Gibbs iterations and look for little variation in the generated distribution over the replications.

Example: A univariate Autoregression

- Consider the following autoregressive model

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + e_t, e_t \tilde{} i.i.d. N(0, \sigma^2),$$

where we assume that the roots of $(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \phi_4 L^4) = 0$ lie outside the unit circle.

- We can write the autoregressive model in matrix notation as

$$Y = X\beta + \varepsilon, \qquad \varepsilon \tilde{} N(0, \sigma^2 I)$$

where $\beta = [\mu,\ \phi_1,\ \phi_2,\ \phi_3,\ \phi_4]'$ and $X = [1,\ y_{t-1},\ y_{t-2},\ y_{t-3},\ y_{t-4}]$.

- Conditional Distributions of β given σ^2

The *prior distribution* of β is $\beta|\sigma^2 \tilde{} N(\beta_0, \Sigma_0)_{I(s(\phi))}$, where $I(s(\phi))$ is an indicator to denote that all the roots of $(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \phi_4 L^4) = 0$ lie outside the unit circle.
The *posterior distribution* of β is $\beta|\sigma^2, Y \tilde{} N(\beta_1, \Sigma_1)_{I(s(\phi))}$,where

$$\begin{aligned} \beta_1 &= (\Sigma_0^{-1} + \sigma^{-2} X'X)^{-1} (\Sigma_0^{-1} \beta_0 + \sigma^{-2} X'Y) \\ \Sigma_1 &= (\Sigma_0^{-1} + \sigma^{-2} X'X)^{-1} \end{aligned}$$

- Conditional Distributions of σ^2 given β

The *Prior* distribution of $\sigma^2|\beta \tilde{} I\Gamma(\frac{v_0}{2}, \frac{\delta_0}{2})$ where v_0 and δ_0 are known and $I\Gamma$ denotes inverted Gamma distribution.
The *posterior distribution* of $\sigma^2|\beta \tilde{} I\Gamma(\frac{v_1}{2}, \frac{\delta_1}{2})$ where
$\delta_1 = \delta_0 + (Y - X\beta)'(Y - X\beta)$ and $v_1 = v_0 + T$.

The Gibbs Sampling procedure consists of the following steps

- Start the iteration of the Gibbs Sampling

-To start the iteration we use an arbitrary starting value $\sigma^2 = \left\{\sigma^2\right\}^0$
-Iterate the following steps $j = L + M$ times

- Conditional on $\sigma^2 = \left\{\sigma^2\right\}^{j-1}$, that is, the value generated in the previous iteration, we generate β^j from the posterior distribution of β.
- Conditional on $\beta = \beta^j$,that is, the value β generated in 1), we generate $\left\{\sigma^2\right\}^j$ from the posterior distribution of σ^2.
- Set $j = j + 1$.

- In generating β we employ rejection sampling to ensure that all the roots are outside the unit circle (we discard the draws that do not satisfy this condition).

As a result of this procedure we generate the following sets of values

$$\beta^1, \beta^2, ..., \beta^{L+M},$$
$$\left\{\sigma^2\right\}^1, \left\{\sigma^2\right\}^2, ..., \left\{\sigma^2\right\}^{L+M}.$$

- We discard the first L generated values to ensure convergence of the Gibbs-Sampler and then use the following M values to make inferences about β and σ^2. The remaining M values provide us with the Joint and the Marginal distribution.

Markov- Switching and Gibbs Sampling

- We have shown that when estimationg M-S models we treat parameters of the model depending on an unobserved sate variable. We typically estimate the models and make inferences on the unobserved variables conditional on the parameters (estimates) of the model.
- The Bayesian approach treats both the parameters of the model and the Markov switching variables as random variables. Then the inference about the states of the economy (denoted as $\widetilde{S}_T = S_1, S_2, ...S_T$) is based on the joint distribution of the sates and the parameters of the model.

- Consider the following model

$$
\begin{aligned}
y_t &= \mu_{S_t} + \varepsilon_t \qquad \varepsilon_t \tilde{} N(0, \sigma^2_{S_t}) \\
\mu_{S_t} &= \mu_0 + \mu_1 S_t, \qquad \mu_1 > 0 \\
\sigma^2_{S_t} &= \sigma^2_0(1 - S_t) + \sigma^2_1 S_t \\
&= \sigma^2_0(1 + h_1 S_t), \qquad h_1 > 0
\end{aligned}
$$

$$
\begin{aligned}
P(S_t &= 0 | S_{t-1} = 0) = q \\
P(S_t &= 1 | S_{t-1} = 1) = p
\end{aligned}
$$

- The Bayesian approach will entail the inference about $T+6$ random variables: $\{S_1, S_2, ..., S_T, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q\}$.

We need to derive the joint posterior distribution

$$\begin{aligned} g(\widetilde{S}_T, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q|\widetilde{y}_T) &= g(\mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q|\widetilde{y}_T, \widetilde{S}_T).g(\widetilde{S}_T|\widetilde{y}_T) \\ &= g(\mu_0, \mu_1, \sigma_0^2, \sigma_1^2|\widetilde{y}_T, \widetilde{S}_T) \\ & \quad g(p, q|\widetilde{y}_T, \widetilde{S}_T).g(\widetilde{S}_T|\widetilde{y}_T) \\ &= g(\mu_0, \mu_1, \sigma_0^2, \sigma_1^2|\widetilde{y}_T, \widetilde{S}_T) \\ & \quad g(p, q|\widetilde{S}_T).g(\widetilde{S}_T|\widetilde{y}_T) \end{aligned}$$

We assume that conditional on $\widetilde{S}_T$, p and q are independent of both other parameters of the model and of the data. Notice that conditional on $\widetilde{S}_T$, the expression $y_t = \mu_{S_t} + \varepsilon_t$, $\varepsilon_t \tilde{} N(0, \sigma_{S_t}^2)$ is simply a regression model with a known dummy.

The Gibbs Sampling estimation procedure

Using arbitrary starting values we repeat the following steps.

1. Generate S_t from $g(S_t|\widetilde{S}_{\neq t}, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q, \widetilde{y}_T)$, or generate the whole block $\widetilde{S}_T$ from $g(\widetilde{S}_T|\mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q, \widetilde{y}_T)$
2. Generate the transition probabilities p and q from $g(p, q|\widetilde{S}_T)$.
3. Generate $\mu_0, \mu_1, \sigma_0^2, \sigma_1^2$, from $g(\mu_0, \mu_1, \sigma_0^2, \sigma_1^2|\widetilde{y}_T, \widetilde{S}_T)$.

- Step 1:Single move Gibbs Sampling, Generate S_t from $g(S_t|\widetilde{S}_{\neq t}, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q, \widetilde{y}_T)$

Suppressing the conditioning in the parameters we can write

$$
\begin{aligned}
g(S_t|\widetilde{S}_{\neq t}, \widetilde{y}_T) &= g(\overbrace{S_t}^{A}|\overbrace{\widetilde{S}_{\neq t}, \widetilde{y}_t}^{C}, \overbrace{y_{t+1}, y_{t+2}, ..., y_T}^{B}) \\
&= \frac{g(\overbrace{S_t}^{A}, \overbrace{y_{t+1}, y_{t+2}, ..., y_T}^{B}|\overbrace{\widetilde{S}_{\neq t}, \widetilde{y}_t}^{C})}{g(\overbrace{y_{t+1}, y_{t+2}, ..., y_T}^{B}|\overbrace{\widetilde{S}_{\neq t}, \widetilde{y}_t}^{C})} \\
\text{since } g(AB|C) &\equiv g(A|BC)g(B|C) \\
&= \frac{g(S_t|\widetilde{S}_{\neq t}, \widetilde{y}_t)g(y_{t+1}, y_{t+2}, ..., y_T|\widetilde{S}_{\neq t}, \widetilde{y}_t)}{g(y_{t+1}, y_{t+2}, ..., y_T|\widetilde{S}_{\neq t}, \widetilde{y}_t)}
\end{aligned}
$$

Since conditional in the state, S_t *and* $y_{t+1}, .., y_T$ are independent.

$$
\begin{aligned}
&= g(S_t|\widetilde{S}_{\neq t}, \widetilde{y}_t) \\
&= g(S_t|\widetilde{S}_{t-1}, S_{t+1}, ..., S_T, \widetilde{y}_{t-1}, y_t)
\end{aligned}
$$

$$g(\overbrace{S_t}^{A}|\overbrace{\widetilde{S}_{t-1},\widetilde{y}_{t-1}}^{C},\overbrace{S_{t+1},...,S_T,y_t}^{B}) = \frac{g(\overbrace{S_t}^{A},\overbrace{S_{t+1},...,S_T,y_t}^{B}|\overbrace{\widetilde{S}_{t-1},\widetilde{y}_{t-1}}^{C})}{g(\overbrace{S_{t+1},...,S_T,y_t}^{B}|\overbrace{\widetilde{S}_{t-1},\widetilde{y}_{t-1}}^{C})}$$
$$\propto g(S_t,S_{t+1},...,S_T,y_t|\widetilde{S}_{t-1},\widetilde{y}_{t-1})$$

$$\begin{aligned} g(S_t,..,S_T,y_t|\widetilde{S}_{t-1},\widetilde{y}_{t-1}) &= g(S_t|\widetilde{S}_{t-1},\widetilde{y}_{t-1})g(S_{t+1},..,S_T,y_t|S_t,\widetilde{S}_{t-1},\widetilde{y}_{t-1}) \\ &= g(S_t|S_{t-1})g(S_{t+1},..,S_T,y_t|S_t,\widetilde{S}_{t-1},\widetilde{y}_{t-1}) \\ &\quad \text{Since is homogeneous Markov.} \end{aligned}$$

Notice that

$$
\begin{aligned}
g(\overbrace{y_t}^{A}, \overbrace{S_{t+1}, .., S_T}^{B} | \overbrace{S_t, \widetilde{S}_{t-1}, \widetilde{y}_{t-1}}^{C}) &= g(\overbrace{y_t}^{A} | \overbrace{S_t, \widetilde{S}_{t-1}, \widetilde{y}_{t-1}}^{C}) \\
& \quad g(\overbrace{S_{t+1}, .., S_T}^{B} | \overbrace{S_t, \widetilde{S}_{t-1}, \widetilde{y}_{t-1}}^{C} \overbrace{y_t}^{A}) \\
&= g(y_t | S_t) g(S_{t+1} | S_t, \widetilde{S}_{t-1}, \widetilde{y}_{t-1}, y_t) \\
& \quad g(S_{t+2}, .., S_T, | S_{t+1}, S_t, \widetilde{S}_{t-1}, \widetilde{y}_{t-1}, y_t) \\
&= g(y_t | S_t) g(S_{t+1} | S_t) g(S_{t+2}, .., S_T | S_{t+1}) \\
&\propto g(y_t | S_t) g(S_{t+1} | S_t)
\end{aligned}
$$

Then, we can write

$$g(S_t|\widetilde{S}_{\neq t}, \widetilde{y}_T) \propto g(S_t|S_{t-1})g(y_t|S_t)g(S_{t+1}|S_t)$$

where $g(S_t|S_{t-1})$ and $g(S_{t+1}|S_t)$ are given by the transition probabilities and

$$g(y_t|S_t) = \frac{1}{\sqrt{2\pi\sigma^2_{s_t}}} \exp\{-\frac{1}{2\sigma^2_{s_t}}(y_t - \mu_{s_t})^2\}$$

We can then calculate

$$P(S_t = j|\widetilde{S}_{\neq t}, \widetilde{y}_T) = \frac{g(S_t = j|\widetilde{S}_{\neq t}, \widetilde{y}_T)}{\sum_{j=0}^{1} g(S_t = j|\widetilde{S}_{\neq t}, \widetilde{y}_T)}$$

We generate S_t using a uniform distribution between 0 and 1. If the generated number is less or equal than $P(S_t = j|\widetilde{S}_{\neq t}, \widetilde{y}_T)$, we set the value of S_t to zero, otherwise we set the value equal to one.

- Step 1):Multimove Gibbs Sampling - Generate $\widetilde{S}_t$ from $g(\widetilde{S}_T|\mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p, q, \widetilde{y}_T)$

Suppressing the conditioning in the parameters we can write

$$
\begin{aligned}
g(\widetilde{S}_T|\widetilde{y}_T) &= g(S_1, S_2, ..., S_T, |\widetilde{y}_T) \\
&= g(S_T, |\widetilde{y}_T) g(S_1, S_2, ..., S_{T-1}, |S_T, \widetilde{y}_T) \\
&= g(S_T, |\widetilde{y}_T) g(S_{T-1}, |S_T, \widetilde{y}_T) g(S_1, ., S_{T-2}, |S_{T-1}, S_T, \widetilde{y}_T) \\
&= g(S_T, |\widetilde{y}_T) g(S_{T-1}, |S_T, \widetilde{y}_T) g(S_{T-2}, |S_{T-1}, S_T, \widetilde{y}_T) .. \\
&\quad .....g(S_1, |S_2, ., S_{T-2}, S_{T-1}, S_T, \widetilde{y}_T) ... \\
(NB &: g(S_{T-1}, |S_T, \widetilde{y}_T) = g(S_{T-1}, |S_T, \widetilde{y}_{T-1}) \\
&\quad \text{since cond on } S_T,\ y_T \text{ adds no info.}) \\
&= g(S_T, |\widetilde{y}_T) g(S_{T-1}, |S_T, \widetilde{y}_{T-1}) \\
&\quad g(S_{T-2}, |S_{T-1}, \widetilde{y}_{T-2}) .. g(S_1, |S_2, y_1) \\
&= g(S_T, |\widetilde{y}_T) \prod_{t=1}^{T-1} g(S_t, |S_{t+1}, \widetilde{y}_t).
\end{aligned}
$$

- The derivation is based on the Markov property that states that to makes inference about S_t conditional on S_{t+1} the variables S_{t+2},....S_T , y_{t+1},....y_T have no information beyond that contained in S_{t+1}.
- Then we proceed in the following way: we first generate $\widetilde{S}_T$ conditional on $\widetilde{y}_T$ and then, for the other values of $t = T-1, t-2, ..., 1$, we generate S_t conditional on y_t and the generated $t+1$.

We can carry out this using the following steps:

Step 1 Run the Hamilton filter to get $g(S_t|\widetilde{y}_t)$. The last iteration of the filter provides $g(S_T|\widetilde{y}_T)$ that is used to generate S_T.

Step 2 Generate S_t conditional on S_{t+1} and $\widetilde{y}_t$, for $t = T-1, t-2, ..., 1$, form $g(S_t|S_{t+1}, \widetilde{y}_t)$ using the fact that

$$\begin{aligned} g(S_t|S_{t+1}, \widetilde{y}_t) &= \frac{g(S_t, S_{t+1}|\widetilde{y}_t)}{g(S_{t+1}|\widetilde{y}_t)} \\ &= \frac{g(S_{t+1}|S_t, \widetilde{y}_t).g(S_t|\widetilde{y}_t)}{g(S_{t+1}|\widetilde{y}_t)} \\ &= \frac{g(S_{t+1}|S_t).g(S_t|\widetilde{y}_t)}{g(S_{t+1}|\widetilde{y}_t)} \\ &\propto g(S_{t+1}|S_t).g(S_t|\widetilde{y}_t) \end{aligned}$$

Then we calculate

$$P(S_t = 1|S_{t+1}, \widetilde{y}_t) = \frac{g(S_{t+1}|S_t = 1)g(S_t = 1|\widetilde{y}_t)}{\sum_{j=0}^{1} g(S_{t+1}|S_t = j)g(S_t = j|\widetilde{y}_t)}$$

We generate S_t using a uniform distribution between 0 and 1. If the generated number is less or equal than $P(S_t = 1|S_{t+1}, \widetilde{y}_t)$, we set the value of S_t to zero, otherwise we set the value equal to one.

- Generating Transition Probabilities p and q, conditional on $\widetilde{S}_T$

Conditional on $\widetilde{S}_T$, p and q are independent of the data set $\widetilde{y}_T$, and the other parameters of the models[2].

- Prior Distribution

$p \tilde{} beta(u_{11}, u_{10})$,
$q \tilde{} beta(u_{00}, u_{01})$,
with $g(p,q) \propto p^{u_{11}-1}(1-p)^{u_{10}-1} q^{u_{00}-1}(1-q)^{u_{01}-1}$, where the $u\prime s$ are known hyper parameters of the priors.

- The likelihood function is given by

$L(p,q|\widetilde{S}_T) = p^{n_{11}}(1-p)^{n_{10}} q^{n_{00}}(1-q)^{n_{01}}$ where n_{ij} refers to the number of transitions from i to j which can be counted from $\widetilde{S}_T$.

$$\begin{aligned} g(z|\alpha_0,\alpha_1) &\propto z^{\alpha_0-1}(1-z)^{\alpha_1-1} \text{ for } 0<z<1 \\ &= 0 \qquad\qquad\qquad\qquad \text{otherwise}, \end{aligned}$$

with $E(z) = \frac{\alpha_0}{\alpha_0+\alpha_1}$ and $Var(z) = \frac{\alpha_0\alpha_1}{(\alpha_0+\alpha_1)^2(\alpha_0+\alpha_1+1)}$

Posterior distribution

$$
\begin{aligned}
g(p,q|\widetilde{S}_T) &= g(p,q)L(p,q|\widetilde{S}_T) \\
&\propto p^{u_{11}-1}(1-p)^{u_{10}-1}q^{u_{00}-1}(1-q)^{u_{01}-1} \\
&\quad p^{n_{11}}(1-p)^{n_{10}}q^{n_{00}}(1-q)^{n_{01}} \\
&= p^{u_{11}+n_{11}-1}(1-p)^{u_{10}+n_{10}-1}q^{u_{00}+n_{00}-1}(1-q)^{u_{01}+n_{01}-1},
\end{aligned}
$$

then,
$p|\widetilde{S}_T \tilde{} beta(u_{11+}n_{11}, u_{10}+n_{10})$,
$q|\widetilde{S}_T \tilde{} beta(u_{00}+n_{00}, u_{01}+n_{01})$.

- Generating μ_0, μ_1,conditional on $\sigma_0^2, \sigma_1^2, \widetilde{y}_T$ and $\widetilde{S}_T$

Given

$$
\begin{aligned}
y_t &= \mu_0 + \mu_1 S_t + \varepsilon_t \qquad \varepsilon_t \tilde{} N(0, \sigma^2_{S_t}) \\
&\text{we can write} \\
y_t^* &= \mu_0 x_{0t} + \mu_1 x_{1t} + v_t \qquad v_t \tilde{} N(0,1) \\
\text{where } y_t^* &= \frac{y_t}{\sigma_{S_t}},\ x_{0t} = \frac{1}{\sigma_{S_t}} \text{ and } x_{1t} = \frac{S_t}{\sigma_{S_t}}
\end{aligned}
$$

- Prior Distribution: We can write the model in matrix notation as

$$Y = X\mu + V, \qquad V \tilde{} N(0, I)$$

then if we assume a normal prior $\mu|\sigma_0^2, \sigma_1^2 \tilde{} N(b_0, B_0)$, where b_0, B_0 are given.

- Posterior distribution, $\mu|\sigma_0^2, \sigma_1^2, \widetilde{S}_T, \widetilde{y}_T \tilde{} N(b_1, B_1)$, where b_1, B_1 are

$$\begin{aligned} b_1 &= (B_0^{-1} + X'X)^{-1}(B_0^{-1}b_0 + X'Y) \\ B_1 &= (B_0^{-1} + X'X)^{-1} \end{aligned}$$

to constrain $\mu_1 > 0$, we discard the draws where this condition is not satisfied.

- Generating σ_0^2, σ_1^2,conditional on $\mu_0, \mu_1, \widetilde{y}_T$ and $\widetilde{S}_T$

Given

$$\begin{aligned} \sigma^2_{S_t} &= \sigma_0^2(1 + h_1 S_t), \qquad h_1 > 0 \\ & \textit{where} \\ \sigma_1^2 &= \sigma_0^2(1 + h_1). \end{aligned}$$

we can first generate σ_0^2 conditional on h_1, and then generate $(1 + h_1)$ conditional on σ_0^2.

- Generating σ_0^2 conditional on h_1

We divide both sides of y_t by $\sqrt{(1+h_1 S_t)}$:

$$\begin{aligned} y_t^{**} &= \mu_0 x_{0t}^* + \mu_1 x_{1t}^* + v_t^* \qquad v_t^* \tilde{} N(0, \sigma_0^2) \\ \text{where } y_t^{**} &= \frac{y_t}{\sqrt{(1+h_1 S_t)}},\ x_{0t}^* = \frac{1}{\sqrt{(1+h_1 S_t)}}, \\ x_{1t}^* &= \frac{S_t}{\sqrt{(1+h_1 S_t)}} \text{ and } v_t^* = \frac{\varepsilon_t}{\sqrt{(1+h_1 S_t)}} \end{aligned}$$

then:
The Prior distribution of $\sigma_0^2|\mu_0, \mu_1, h_1 \tilde{} I\Gamma(\frac{v_0}{2}, \frac{\delta_0}{2})$ where v_0 and δ_0 are known and $I\Gamma$ denotes inverted Gamma distribution.
The posterior distribution of $\sigma_0^2|\mu_0, \mu_1, h_1, \widetilde{S}_T, \widetilde{y}_T \tilde{} I\Gamma(\frac{v_1}{2}, \frac{\delta_1}{2})$ *where*
$\delta_1 = \delta_0 + \sum_{t=1}^{T} (y_t^{**} - \mu_0 x_{0t}^* - \mu_1 x_{1t}^*)^2$ and $v_1 = v_0 + T$.

- Generating h_1 conditional on σ_0^2

We divide both sides of y_t by σ_0 :

$$\begin{aligned} y_t^{***} &= \mu_0 x_{0t}^{**} + \mu_1 x_{1t}^{**} + v_t^{**} \qquad v_t^{**} \sim N(0, (1 + h_1 S_t)) \\ \text{where } y_t^{***} &= \frac{y_t}{\sigma_0},\ x_{0t}^{**} = \frac{1}{\sigma_0},\ x_{1t}^{**} = \frac{S_t}{\sigma_0} \text{ and } v_t^{**} = \frac{\varepsilon_t}{\sigma_0} \end{aligned}$$

then:

- The Prior distribution of $h_1|\mu_0, \mu_1, \sigma_0^2 \sim I\Gamma(\frac{v_2}{2}, \frac{\delta_2}{2})$ where v_2 and δ_2 are known and $I\Gamma$ denotes inverted Gamma distribution.
- The posterior distribution of $h_1|\mu_0, \mu_1, \sigma_0^2, \widetilde{S}_T, \widetilde{y}_T \sim I\Gamma(\frac{v_3}{2}, \frac{\delta_3}{2})$ *where* $\delta_3 = \delta_2 + \sum_{t=1}^{N_1} (y_t^{***} - \mu_0 x_{0t}^{**} - \mu_1 x_{1t}^{**})^2$ and $v_3 = v_2 + T$, where N_1 is the number of times $S_t = 1$.