Nonlinear times series modelling

Kevin Kotzé

Contents

Introduction

- Econometrics: use available information to describe relationships
- Time series usually extend over quite a long period of time
- Over such a period we often incur certain changes that influence the behaviour of the DGP (e.g. expansion-recession, etc.)
- Regime switching models incorporate the dynamic state dependent behaviour of economic or financial variables

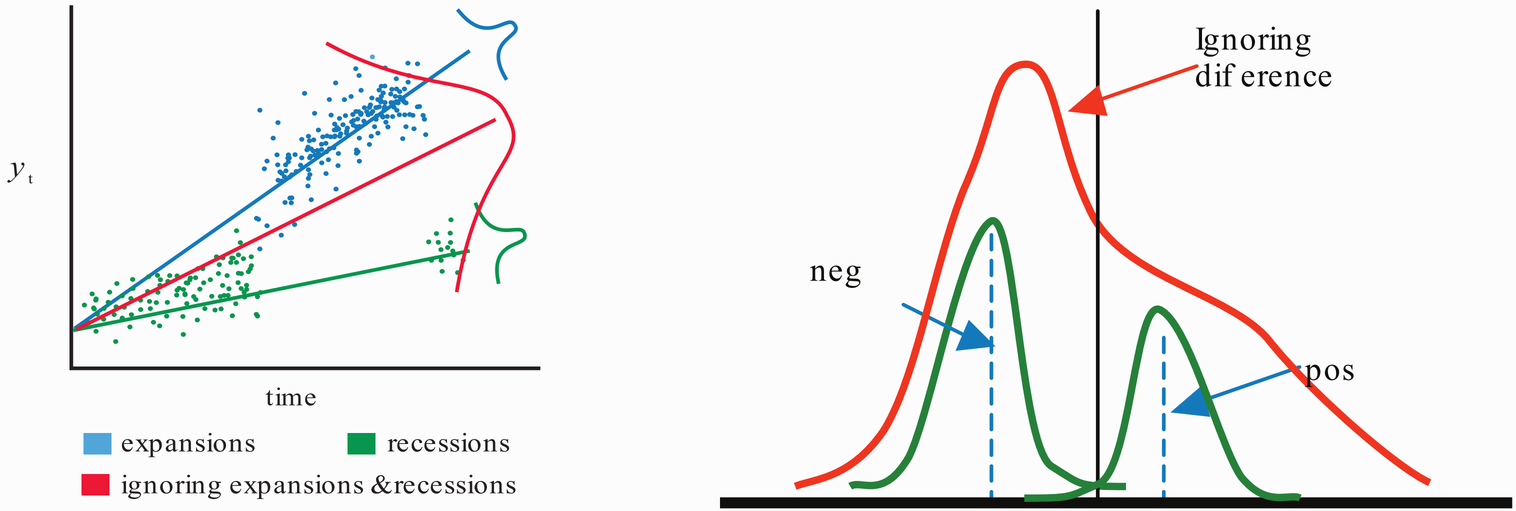


Figure : Regime switching DGPs

Further examples

- Further examples of such behaviour occur in stock markets
- Skewness (large negative returns are more common than large positive returns) and
- Kurtosis (large absolute returns are frequently observed)
- Other uses:
  - determination of the actual change
  - multiple equilibria
- Other nonlinear models (GAR, bilinear,etc)

Dummy variable models

- Assume the sample can be split into separate groups (regimes)

- Parameters are constant within the groups (regimes) but differ between groups
  - There is no transition period

- Identification of the groups (regimes) is known with certainty in advance

- Regime switching process is deterministic

Dummy variable models

- Dummy removes that part of the residual that was generated as a result of a sudden change
- Facilitates more accurate parameter estimation
- Chow break test used to determine if break is significant
    - Split sample into groups and test null hypothesis of constant coefficients in subgroups
    - Could also use CUSUM test
    - Tests for possible parameter instability in sample

Basic Regime Switching

- Regime is described by a stochastic process
- Future regimes are not know with certainty
  - regime is determined by an observable variable (past & present known)
  - regime is determined by an unobserved stochastic process (assign probabilities to regime)
- Simple specification

$$y_t = \begin{cases} \phi_{0,1} + \phi_{1,1} x_t + \varepsilon_t & \text{in regime one} \\ \phi_{0,2} + \phi_{1,2} x_t + \varepsilon_t & \text{in regime two} \end{cases}$$

- where $\varepsilon_t \sim (0, \sigma_i^2)$ in regime $i, i = 1, 2$

Basic Regime Switching

- In contrast with dummy variable models:
  - Each regime is modeled explicitly
  - May have different errors for each regime % may be of relevance when interpreting the individual shocks from each regime
  - Important when interpreting the model parameters
  - Relevant for identifying the current regime

Threshold Autoregressive (TAR) model

- Regime can be described by observable variable, q_t , relative to a threshold, c

$$y_t = \begin{cases} \phi_{0,1} + \phi_{1,1} x_t + \varepsilon_t & \text{if } q_t \leq c \\ \phi_{0,2} + \phi_{1,2} x_t + \varepsilon_t & \text{if } q_t > c \end{cases}$$

- where $\varepsilon_t \sim (0, \sigma_i^2)$ in regime $i, i = 1, 2$

Threshold Autoregressive (TAR) model

- Consider Shen & Hakes (1995)
  - reaction function of the Taiwanese central bank
  - inflation determines which regime they are in
  - targeted inflation rate provides the threshold
  - F-test for nonlinearity
- no inflation: pursue output growth and low inflation
- low inflation: pursue output growth (with no response to inflation)
- moderate & high inflation: pursue only inflation and not output growth

SETAR models

- Self Extracting Threshold Autoregressive models
- Observable variable, q_t , is a lagged value of the series itself
- Hence, for $AR(1)$ SETAR model;

$$y_t = \begin{cases} \phi_{1,1} y_{t-1} + \varepsilon_t & \text{if } y_{t-1} \leq c \\ \phi_{1,2} y_{t-1} + \varepsilon_t & \text{if } y_{t-1} > c \end{cases}$$

- where $\varepsilon_t \sim (0, \sigma_i^2)$ in regime $i, i = 1, 2$
- such that ε_{1t} and ε_{2t} are responsible for the regime switching

SETAR models

- Alternative representation
  - If $\sigma_1^2 = \sigma_2^2 = \sigma^2$, then the model may be written as;

$$
\begin{aligned}
y_t = (\phi_{0,1} + \phi_{1,1} y_{t-1})(1 - I\left[y_{t-1} > c\right]) \ldots \\
\ldots + (\phi_{0,2} + \phi_{1,2} y_{t-1})(I\left[y_{t-1} > c\right]) + \varepsilon_t
\end{aligned}
$$

- where $I[\cdot]$ is an indicator function with $I[\cdot] = 1$ if event 1 occurs and $I[\cdot] = 0$ otherwise

STAR models

- Smooth Transition Autoregressive models
- More gradual transition between regimes
- Weight series with a continuous (logistic) function
- Where γ is the smoothing parameter, $\gamma > 0$

$$G\left[q_t; \gamma, c\right]$$

- Changes smoothly from 0 to 1 as q_t increases
- A popular choice for the transition mechanism is the logistic function, such that;

$$G\left[q_t; \gamma, c\right] = \frac{1}{1 + \exp(-\gamma[q_t - c])}$$

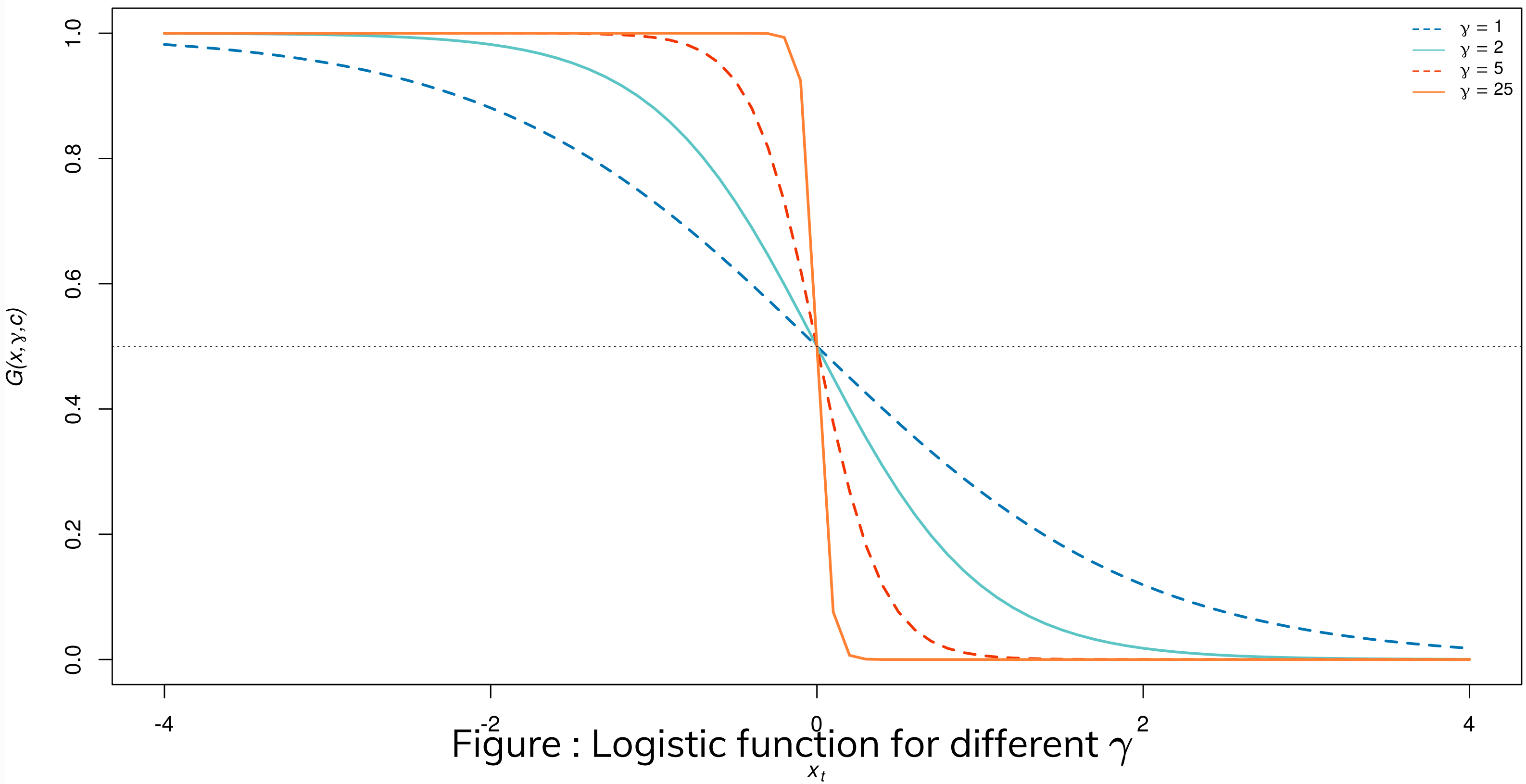


Figure : Logistic function for different γ

- As $\gamma \to \infty$, STAR model represents a TAR model
- As $\gamma \to 0$, STAR model represents a linear model

STAR models

- The resulting expression for the model is;

$$y_t = \phi_1 x_t(1 - G\left[q_t; \gamma, c\right]) + \phi_2 x_t(G\left[q_t; \gamma, c\right]) + \varepsilon_t$$

- where $\varepsilon_t \sim (0, \sigma^2)$ and γ is the smoothing parameter in the continuous function

STAR: Parameter estimation

- Nonlinear least squares is often used to estimate parameters
- Optimization technique that seeks to minimize an objective function with the aid of an iterative search process
  - Consider the standard linear model the estimated parameters and residual sum of squares are calculated as:

$$\begin{aligned} y_t &= E(y_t|\mathbf{x}_t) + \varepsilon_t \\ &= \phi_1 x_{t1} + \phi_2 x_{t2} + \ldots + \phi_p x_{tp} + \varepsilon_t \end{aligned}$$

STAR: Parameter estimation

- The estimated parameters and residual sum of squares are calculated as:

$$\hat{\phi} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}$$
$$S(\hat{\phi}) = \sum_{t=1}^{n} \left[y_t - \hat{\phi}_1 x_{t1} - \hat{\phi}_2 x_{t2} - \ldots - \hat{\phi}_p x_{tp} \right]^2$$

- Least squares estimation involves identifying the parameter values that minimize the sum of square residuals

STAR: Parameter estimation

- Could also have found these parameters with an iterative search technique:

$$S(\hat{\beta}) = \sum_{t=1}^{n} \Big[y_t - \tilde{\phi}_1 x_{t1} - \tilde{\phi}_2 x_{t2} - \ldots - \tilde{\phi}_p x_{tp} \ldots$$

$$\ldots - (\hat{\phi}_1 - \tilde{\phi}_1) x_{t1} - (\hat{\phi}_2 - \tilde{\phi}_2) x_{t2} - \ldots - (\hat{\phi}_p - \tilde{\phi}_p) x_{tp} \Big]^2$$

- where the initial guess for the value of the coefficients in matrix ϕ is expressed as $\tilde{\phi}$

STAR: Parameter estimation

- Why is this means of parameter estimation so useful?
  - Parameters need not only include estimates for ϕ
  - In the STAR model the parameters could include $\psi = (\phi; G[q_t; \gamma, c])$

- Such that

$$y_t = f(\mathbf{x_t}, \psi) + \varepsilon_t \qquad t = 1, 2, \ldots, n$$

- where $\psi = (\psi_1, \psi_2, \ldots, \psi_p)$ represents the parameters in the nonlinear regression, including $\psi = (\phi; G[q_t; \gamma, c])$

STAR: Testing - nonlinearity

- Before estimating a regime switching model test for nonlinearity

- Unfortunately there is no simple test which indicates the form of nonlinearity
  - Likelihood Ratio (LR) test: based on the loss of log-likelihood following the imposition of certain restrictions (i.e. linearity)

- Requires estimates for both linear and nonlinear models

- Large test statistic relative to critical value from χ^2 distribution = reject null
  - Some researchers make use of LM type tests
  - No test is going to tell you which form of nonlinearity is the correct *a priori*

STAR: Testing - diagnostics

- Not all test statistics of linear models are applicable
- Test for serial correlation in each regime

$$\begin{aligned} y_t = (\phi_{0,1} + \phi_{1,1} y_{t-1})(1 - I\left[y_{t-1} > c\right]) \ldots \\ \ldots + (\phi_{0,2} + \phi_{1,2} y_{t-1})(I\left[y_{t-1} > c\right]) + \varepsilon_t \end{aligned}$$

- Could use some form of LR test but this would be time intensive
- Would also need to test for how many regimes to include

STAR: Testing - diagnostics

- Testing for remaining nonlinearity:
  - LM statistic testing the null a 2 regime STAR model capture the nonlinearity (i.e. against the alternative of whether a 3 regime STAR)

- Testing for parameter constancy:
  - investigate whether we would need to include time varying parameters
  - test hypothesis $\gamma_2 = 0$ against the alternative of smoothly changing parameters

- Very few really good in-sample tests - hence most people use extensive out-of-sample tests

STAR example

- Apply a STAR model to describe USDZAR exchange rate:
  - reject the null of linearity using LM test - but it does not tell us what form of nonlinearity to use
  - estimate a two regime STAR model for an $AR(3)$ process
  - threshold variable is the average exchange rate for the last 4 weeks
  - uses NLS optimization procedure to find ϕ and γ parameters
  - starting values are determined with simplified grid search technique

STAR: Example

	Coef	Std.Err	t-val	prob
ϕ 1.1	0.42437	0.20515	2.0686	0.03859
ϕ 1.2	0.47478	0.60845	0.7803	0.4352
ϕ 2.1	-0.22728	3.63697	-0.0625	0.95017
ϕ 2.2	-2.31202	26.49029	-0.0873	0.93045
ϕ 2.3	1.0281	9.9477	0.1034	0.91769
γ	8.61485	5.52041	1.5605	0.11863
thresh	0.15996	1.04446	0.1532	0.87828

AIC = -748\ p = 0.0013 (AR versus STAR)

Specification for regime switching models

- Granger (1993):

Building models of nonlinear relationships are inherently more difficult than linear ones. There are *more possibilities*, many *more parameters* and thus *more mistakes* can be made. It is suggested that a strategy be applied when attempting such modelling involving *testing for linearity*, considering just a *few model types* of *parsimonious* form and performing *post-sample evaluation* of the models compared to a linear one. The strategy proposed is a 'simple-to-general' one and the application of a heteroscedasticity correction is not recommended.

Specification procedures for regime switching models

- Specify a linear model to describe y_t in terms of x_t
- Test the null hypothesis of linearity against the alternative of TAR, STAR, MSW, or ANN nonlinearity
  - portmanteau, RESET and Macleod-Li tests
  - testing the null of linearity in LM type tests
  - existence of unidentified nuisance parameters $(c$ and $\gamma)$ - simulation
- Estimate the parameters in the selected model
- Evaluate the model using diagnostic tests
  - LM type tests: serial correlation, parameter consistency, heteroscedasticity, omitted variables
- Modify the model if necessary
- Use the model for descriptive or forecasting purposes

Markov Switching Models

- Not able to observe reliable variable that we could use as the regime indicator, q_t
- Regime at t derived from unobserved process, S_t
- S_t represents the probability of being in a certain state at a point of time

$$y_t = \begin{cases} \phi_{0,1} + \phi_{1,1} x_t + \varepsilon_t & \text{if } S_t = 0 \\ \phi_{0,2} + \phi_{1,2} x_t + \varepsilon_t & \text{if } S_t = 1 \end{cases}$$

- where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2_{S_t})$ in regime $i, i = 1, 2$

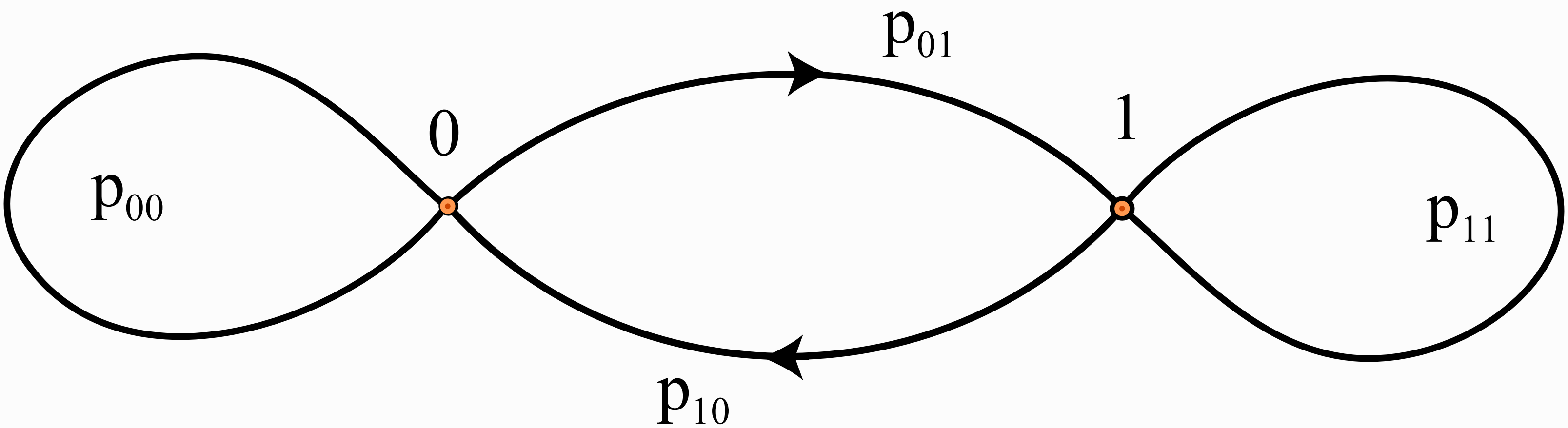


Figure : Markov Chain

Markov Switching Models

- Unobserved process is first order Markov process
- Current regime only depends on previous regime and p_{ij}
- System will either be in regime 0 or regime 1
- At t there is a probability p_{ij} that the system, if in regime i will change to regime j (where $i, j = 1, 0$)

Fixed transition probabilities

- Hence;

$$
\begin{aligned}
P(s_t = 0|s_{t-1} = 0) &= p_{00} \\
P(s_t = 1|s_{t-1} = 0) &= p_{01} = 1 - p_{00} \\
P(s_t = 0|s_{t-1} = 1) &= p_{10} = 1 - p_{11} \\
P(s_t = 1|s_{t-1} = 1) &= p_{11}
\end{aligned}
$$

- where $p_{10} + p_{11} = 1$ and $p_{00} + p_{01} = 1$
- If $p_{10} = p_{01} = 1$ and $p_{00} = p_{11} = 0$, then the system will be in continuously change
- If $p_{10} = p_{01} = 0$ and $p_{00} = p_{11} = 1$, then the system will never change out of one regime
- Conditional states!

Steady states

- If all the probabilities are nonzero then the system will approach a stable point where;

$$P[S_0] = \frac{1 - p_{11}}{2 - p_{00} - p_{11}} \quad \text{or} \quad \frac{p_{10}}{p_{01} + p_{10}}$$
$$P[S_1] = \frac{1 - p_{00}}{2 - p_{00} - p_{11}} \quad \text{or} \quad \frac{p_{01}}{p_{01} + p_{10}}$$

- These steady state probabilities describe the unconditional probabilities of being in each regime at a point in time
- Although the expected duration of being in a particular regime can differ, the probabilities are forced to be constant over time
- Hence - fixed transition probability MSW

Time varying transition probabilities

- May want transitional probabilities to vary over time to include a richer information set. For example,
  - economy in robust recovery is less likely to fall into a recession
- Make p_{ij} a function of duration or a function of another variable
- Hence;

$$
\begin{aligned}
P(s_t = 0 | s_{t-1} = 0, \psi_t) = p_{00}(\psi_t) \\
P(s_t = 1 | s_{t-1} = 0, \psi_t) = p_{01}(\psi_t) \\
P(s_t = 0 | s_{t-1} = 1, \psi_t) = p_{10}(\psi_t) \\
P(s_t = 1 | s_{t-1} = 1, \psi_t) = p_{11}(\psi_t)
\end{aligned}
$$

- where Ω_t is the information set for evolution of the unobserved regime

Varying TVTP

- Could allow for transitional probabilities to be dependent on the value of an exogenous variable, z_t, in a logistic function:

$$p_{00} = P(S - t = 0 | S_{t-1} = 0) = \frac{\exp(\alpha_0 + \beta_0 z_t)}{(1 + \exp(\alpha_0 + \beta_0 z_t)}$$

$$p_{11} = P(S - t = 1 | S_{t-1} = 1) = \frac{\exp(\alpha_1 + \beta_1 z_t)}{(1 + \exp(\alpha_1 + \beta_1 z_t)}$$

Parameter estimation

- Procedure for the Markov Switching model is non-standard since it seeks to obtain;
  - estimates of the parameters in the different regimes
  - estimates for the probability of transition over time
  - estimates for the probability of being in a particular state at a period of time

- Draw observed variable y_t from distribution conditional on the discrete random variable S_t to obtain;

$$f(y_t|S_t, \Omega_{t-1})$$

- Assume the unobserved process is generated by a probability distribution

$$f(y_t|S_t, \Omega_{t-1}) = P[S_t = j|\Omega_{t-1}]$$

Parameter estimation

- Conditional probability from joint probability;

$$P(A \text{ and } B) = P(A|B) \cdot P(B)$$

- To determine the probability that both $S_t = j$ and y_t falls within a predetermined interval, conditional probability is given;

$$\begin{aligned} f(y_t, S_t|\Omega_{t-1}) &= f(y_t|S_t, \Omega_{t-1}) \cdot f(S_t|\Omega_{t-1}) \\ &= f(y_t|S_t, \Omega_{t-1}) \cdot P[S_t = j|\Omega_{t-1}] \end{aligned}$$

- where Ω_{t-1} refers to the information set to $t - 1$ and $j = 0, 1, \ldots $

Hamilton Filter (1989)

- Algorithm responsible for calculating the probability that the process is in regime j at time t given;
  - all observations up to time $t-1$ (i.e. the forecast)
  - all observations up to time t (i.e. the inference)
  - all observations in the entire sample (i.e. the smoothed inference)

- Works similar to the Kalman filter, which may be used to produce values for the unobserved process

Basic procedure

- The following iterative procedure is suggested:
  - obtain starting values for the model parameters
  - compute the smoothed regime probabilities with the aid of the procedures specified in the forecast & inference sections
  - combine these estimates with the initial estimates of the transition probabilities to obtain new estimates for the transition probabilities working backwards from n to 1
  - calculate values for the remaining ϕ parameters
  - iterating this procedure renders a new set of estimates until convergence occurs

Forecasting

- There are no shocks in the out-of-sample period
- Could assume that economy/state persists over the out-of-sample period
- Could generate forecasts under each regime and combine them with the transition probabilities

Applications of Markov Switching

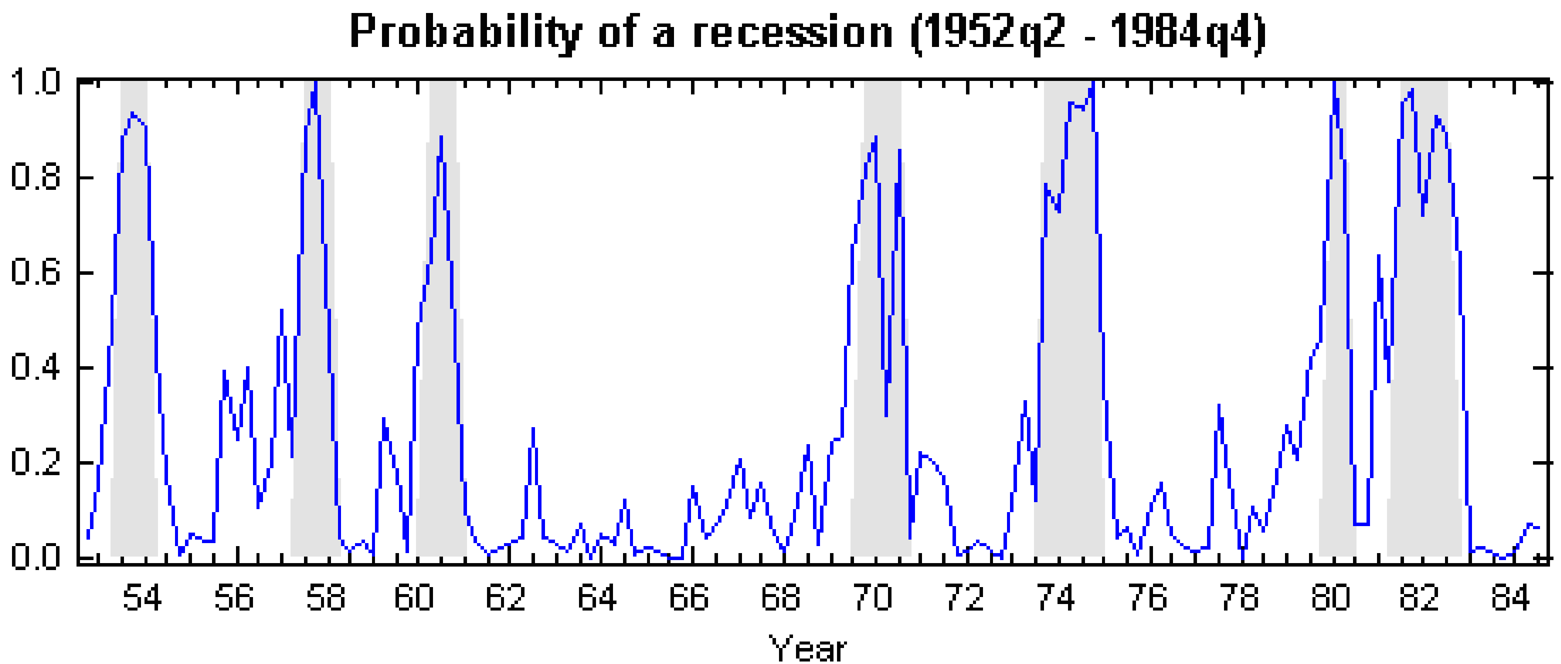
- Business cycle analysis
  - determination of turning points
  - determination of length of business cycle
  - forecasting turning points
- Appreciation and depreciation regimes in exchange rates
- Different regimes in volatility
- Different regimes in interest rates
- Different political regimes and other institutional characteristics

Hamilton's Markov Switching model

- Growth in U.S. Real GNP follows an $AR(4)$ process
- Two state Markov switching (expansion & recession)
- Hence;

$$
\begin{aligned}
(y_t - \mu_{S_t}) = \phi_1(y_{t-1} - \mu_{S_{t-1}}) + \phi_2(y_{t-2} - \mu_{S_{t-2}}) \dots \\
\dots + \phi_3(y_{t-3} - \mu_{S_{t-3}}) + \phi_4(y_{t-4} - \mu_{S_{t-4}}) + \varepsilon_t
\end{aligned}
$$

- where $\varepsilon_t = \text{i. i. d.}\mathcal{N}(0, \sigma^2_{S_t})$



- Where the shaded areas represent the NBER business cycles, this model was able to replicate the business cycles relatively well for the given sample period

Moolman's Markov Switching model

- Moolman (2004) derives a model with TVTP, which are influenced by yield spreads
- Once again growth in S.A. Real GDP follows an $AR(4)$ process
- Two state Markov switching (expansion & recession)
- Hence;

$$
\begin{aligned}
y_t = \mu_2(1 - S_t) + \mu_1 S_t + \phi_1(y_{t-1} - (\mu_2(1 - S_{t-1})) + \mu_{S_{t-1}}) + \ldots \\
\ldots + \phi_4(y_{t-4} - (\mu_2(1 - S_{t-1})) + \mu_{S_{t-4}}) + \varepsilon_t
\end{aligned}
$$

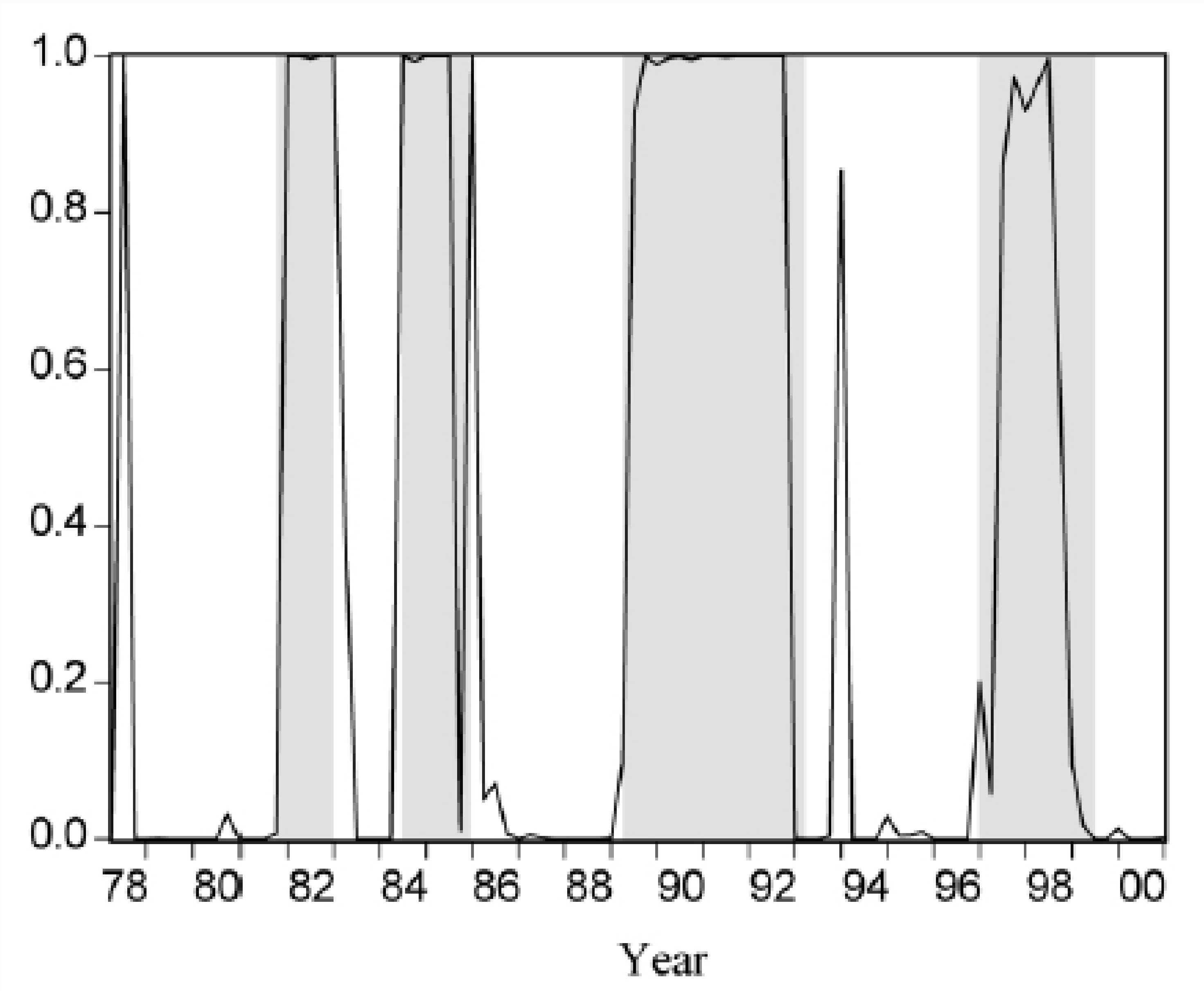


Figure : Moolman's Markov Switching model

Engle's Markov Switching model

- Engle and Hamilton (1990) showed that the US Dollar exchange rate appears to follow long swings as it drifts upwards (downwards) for a considerable period of time

- Engle (1994) showed that for the exchange rates of 18 different countries;
  - MSW model outperforms random walk models on in-sample testing
  - Does not outperform the random walk model or the forward exchange rate in its out-of-sample forecasting ability

- Also suggested that the model picks up a change in regime fairly early on, however, this feature seems to be dependent upon the persistence of a regime

Artificial Neural Network models

- Often regarded as flexible non-parametric models that can approximate any nonlinear function arbitrarily closely
- Could also be specified as flexible regime switching models and could be interpreted as such
- Drawback:
  - parameters are impossible to interpret, hence, only used for pattern recognition and forecasting
  - superior in sample fit could be the result of modelling irregular / unpredictable noise

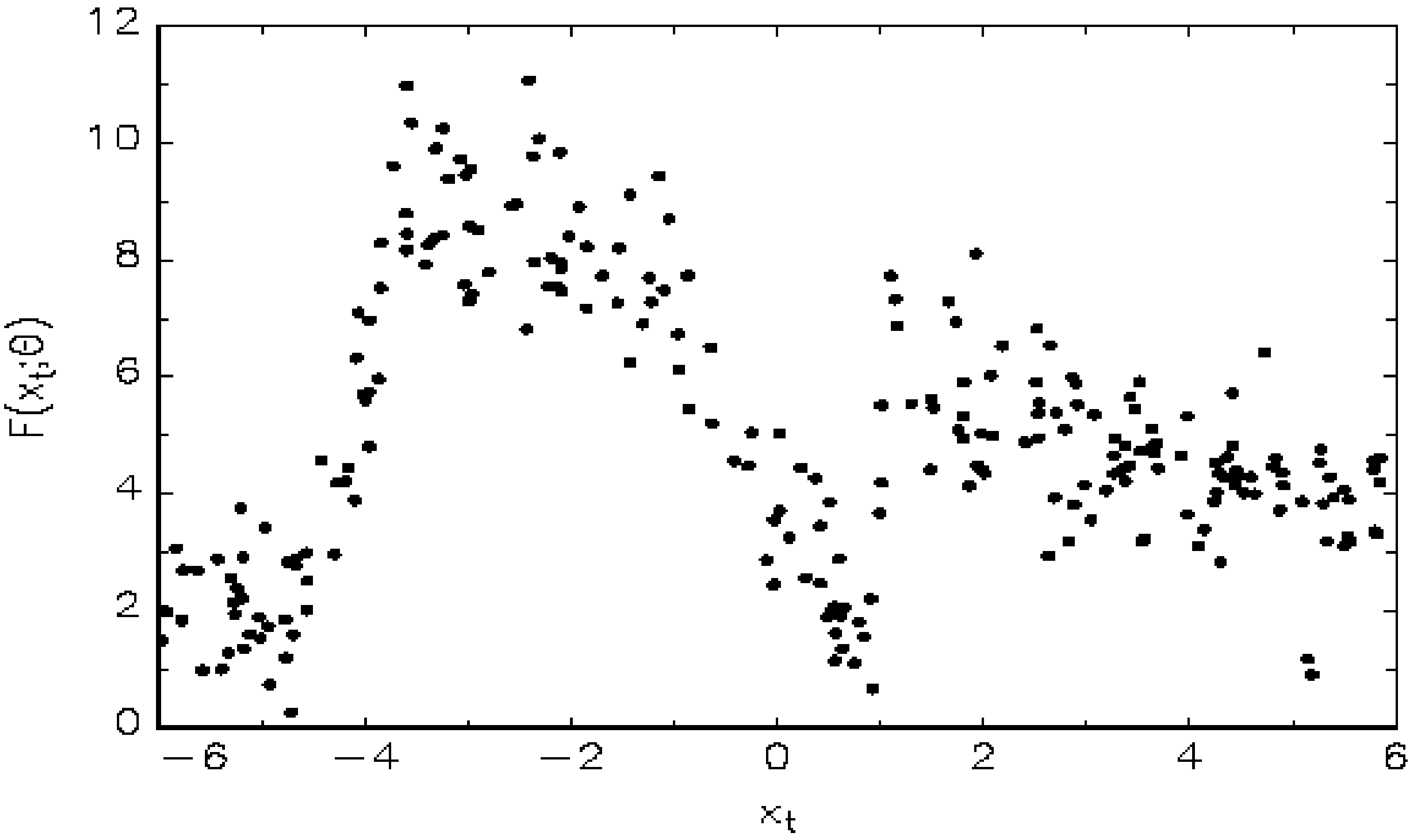


Figure : Scatter plot for hypothetical relationship, x_t and y_t

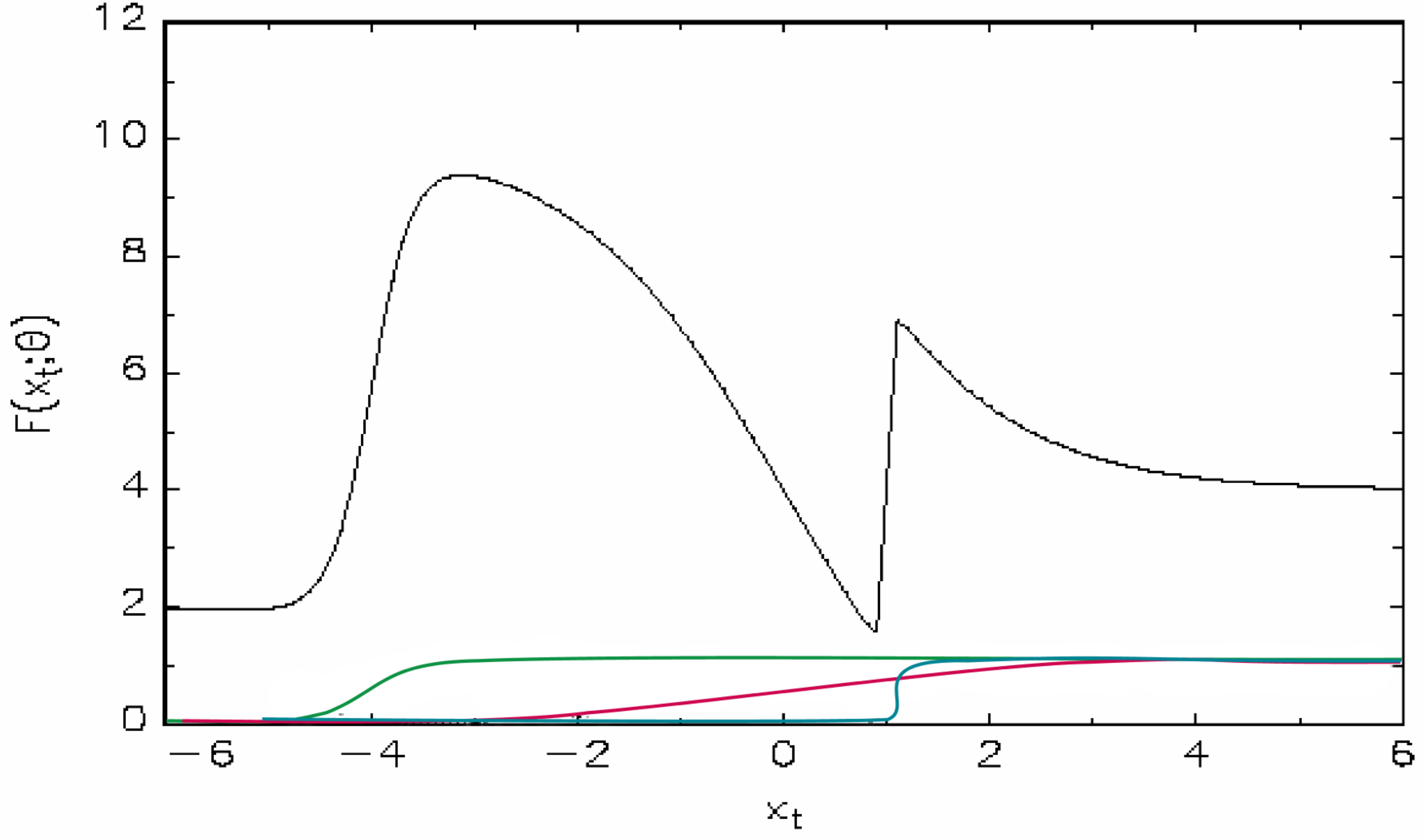


Figure : ANN model for time series with $q = 3$

Estimation

- The idea is similar to a STAR model
- For example, single hidden-layer feedforward ANN takes the form;

$$y_t = \phi_0 + \sum_{j=1}^{1} \beta_j G\left(\gamma_j \left(x_t - c_j\right)\right) + \varepsilon_t$$

- Where the logistic function is used for smoothing $G(\cdot)$;

$$G(z) = \frac{1}{1 + \exp(-z)}$$

Estimation

- Assume parameters $(c_j, j = 1, \ldots, q)$ follow, $c_1 \leq c_2 \leq \ldots \leq c_q$
- Then;

$$\hat{g}(x_t) = \begin{cases} \phi_0 & \text{if } x_t \leq c_1 \\ \phi_0 + \beta_1 & \text{if } c_1 \leq x_t \leq c_2 \\ \phi_0 + \beta_1 + \beta_2 & \text{if } c_2 \leq x_t \leq c_3 \\ \vdots & \\ \phi_0 + \beta_1 + \beta_2 + \ldots + \beta_q & \text{if } c_q < x_t \end{cases}$$

Difference to STAR

- Although looks similar to a STAR model it is different;
- STAR: regime is usually determined by 1 lagged value of y_t,
- ANN: regime normally considers p lagged values of y_t
- STAR: each regime has its own intercept
- ANN: only one intercept is used
- An important difference is ANN models normally use more than one logistic function, which gives it the ability to approximate any nonlinear model arbitrarily closely

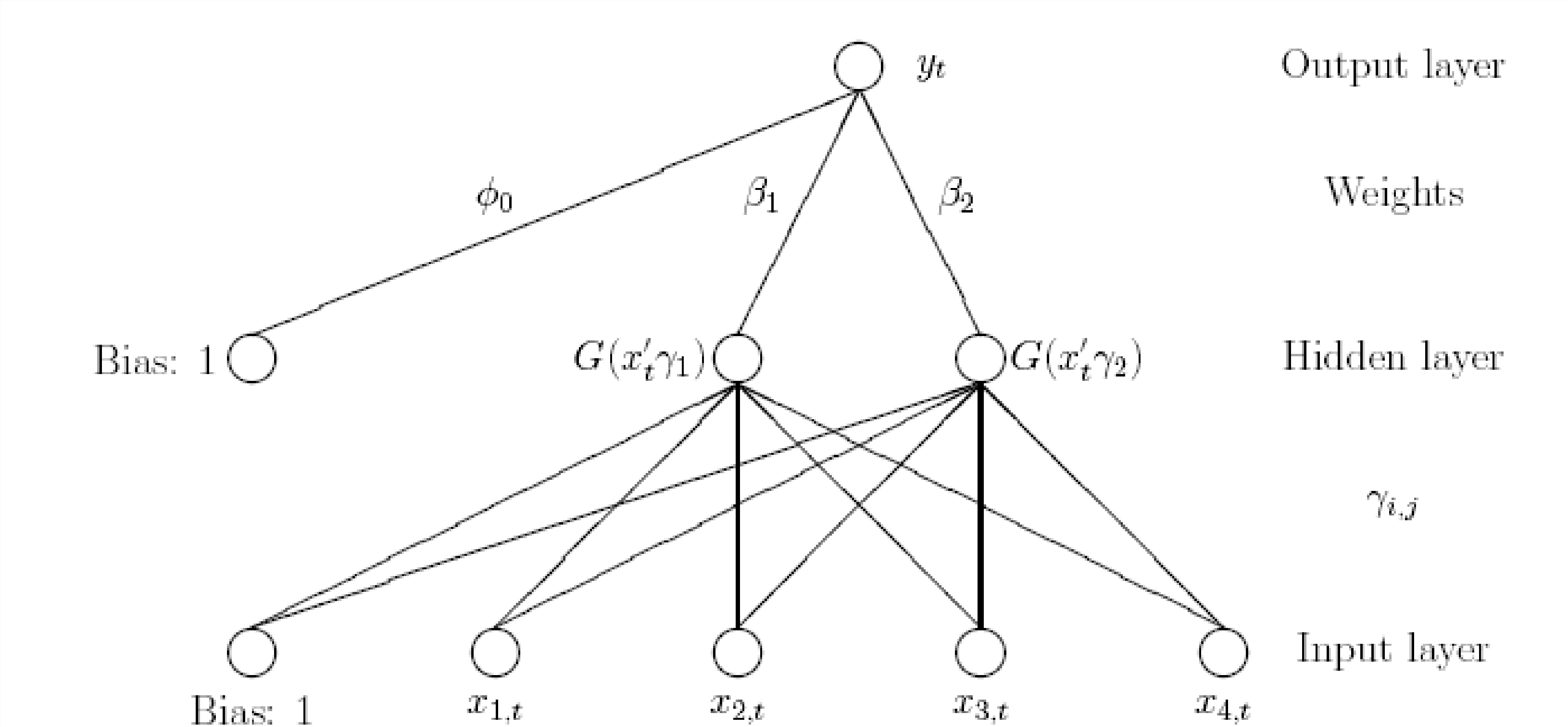


Figure : Traditional (alternative) nomenclature

Testing

- Evaluate the model testing in-sample fit
- Conduct various misspecification tests for remaining nonlinearity and parameter constancy
- Conduct residual diagnostic tests for serial correlation, heteroscedasticity and normality
- Conduct stringent out-of-sample testing - Granger's (1993)