APPLIED ECONOMETRIC TIME SERIES 4RD ED. WALTER ENDERS

Chapter 7

WALTER ENDERS, UNIVERSITY OF ALABAMA

LINEAR VERSUS NONLINEAR ADJUSTMENT

- On a long automobile trip to a new location, you might take along a road atlas. … For most trips, such a linear approximation is extremely useful. Try to envision the nuisance of a nonlinear road atlas.

- For other types of trips, the linearity assumption is clearly inappropriate. It would be disastrous for NASA to use a flat map of the earth to plan the trajectory of a rocket launch.

- Similarly, the assumption that economic processes are linear can provide useful approximations to the actual time-paths of economic variables.

  – Nevertheless, policy makers could make a serious error if they ignore the empirical evidence that unemployment increases more sharply than it decreases.

The Use of Nonlinear Models

- It is now generally agreed that linear econometric models do not capture the dynamic relationships present in many economic time-series.
  - The observation that firms are more apt to raise than to lower prices is a key feature of many macroeconomic models.
  - Neftci (1984), Falk(1986), DeLong and Summers (1988), Granger and Lee (1989), and Teräsvirta and Anderson (1992) establish the result that many real variables display non-linear adjustment over the course of the business cycle.
  - In several papers, Enders and Sandler model many terrorist incident series as nonlinear.
- However, adopting an incorrect non-linear specification may be more problematic than simply ignoring the non-linear structure in the data. It is not surprising, therefore, that non-linear model selection is an important area of current research.
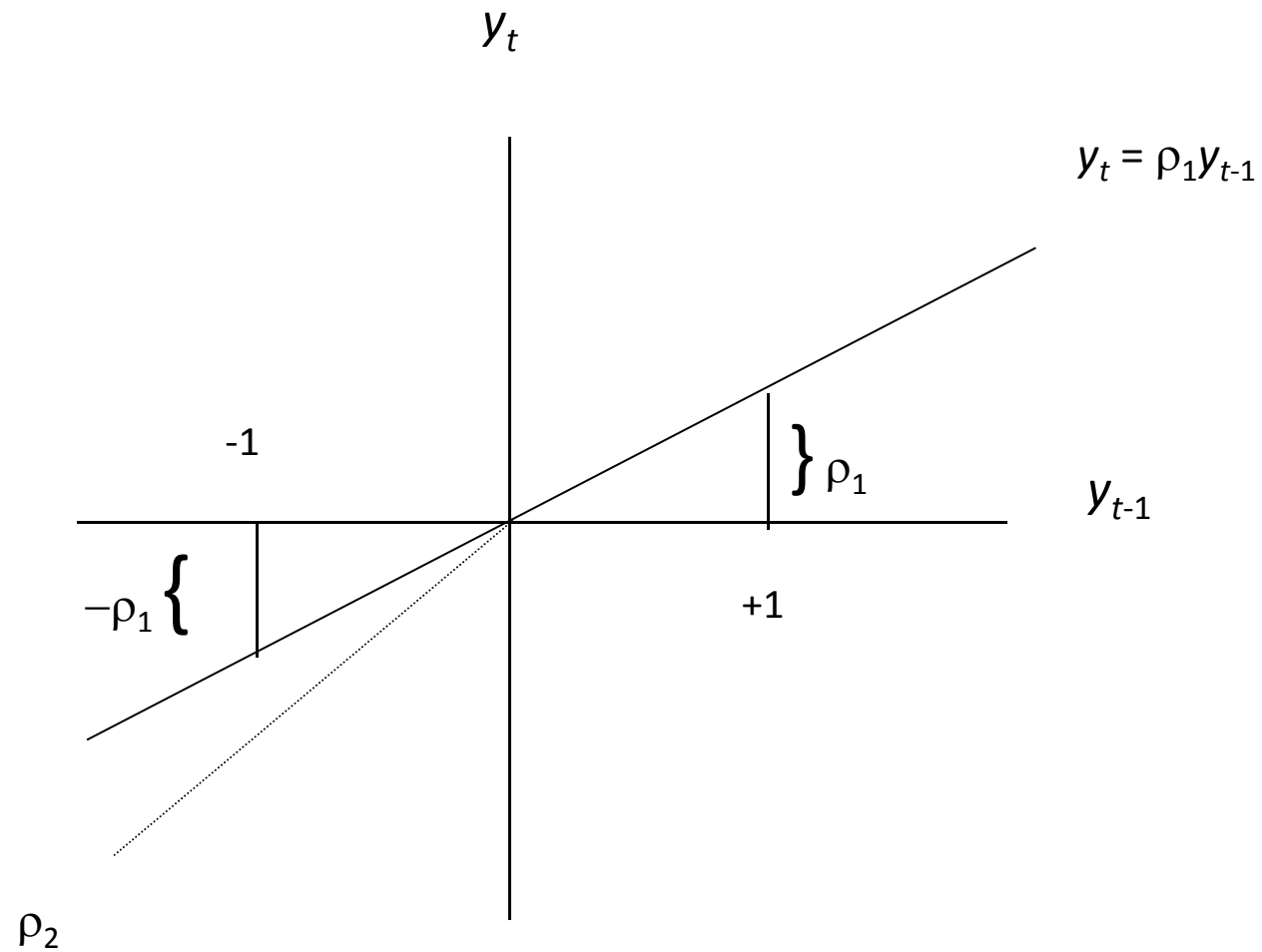
The Interest Rate Spread

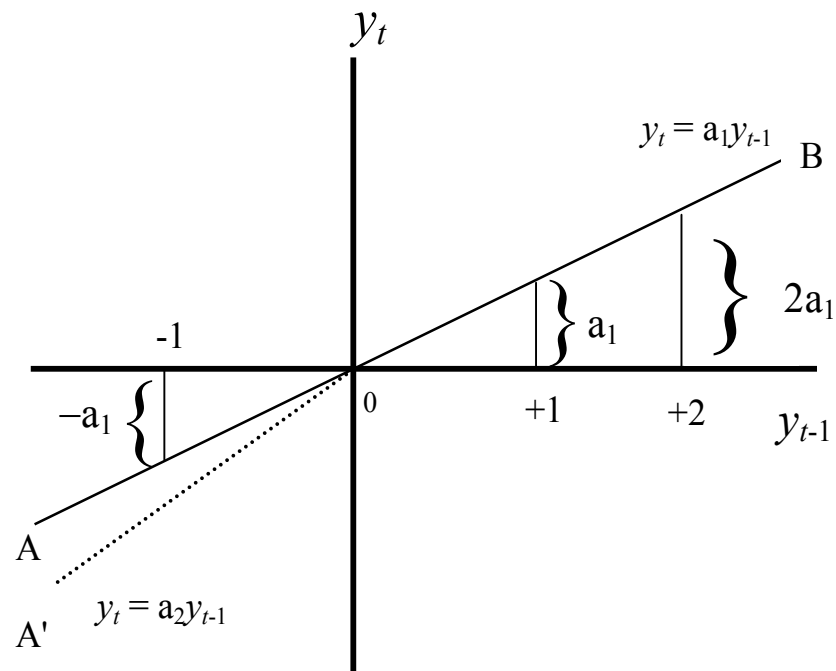- There is evidence that interest rate spreads (s_t) display a nonlinear adjustment pattern.

$$s_t = \begin{cases} \overline{s} + a_1(s_{t-1} - \overline{s}) + \varepsilon_{1t} & \text{when } s_{t-1} > \overline{s} \\ \overline{s} + a_2(s_{t-1} - \overline{s}) + \varepsilon_{2t} & \text{when } s_{t-1} \leq \overline{s} \end{cases}$$

- As long as $|a_2| > |a_1|$, periods when $s_{t-1} <$ will tend to be more persistent than other periods.
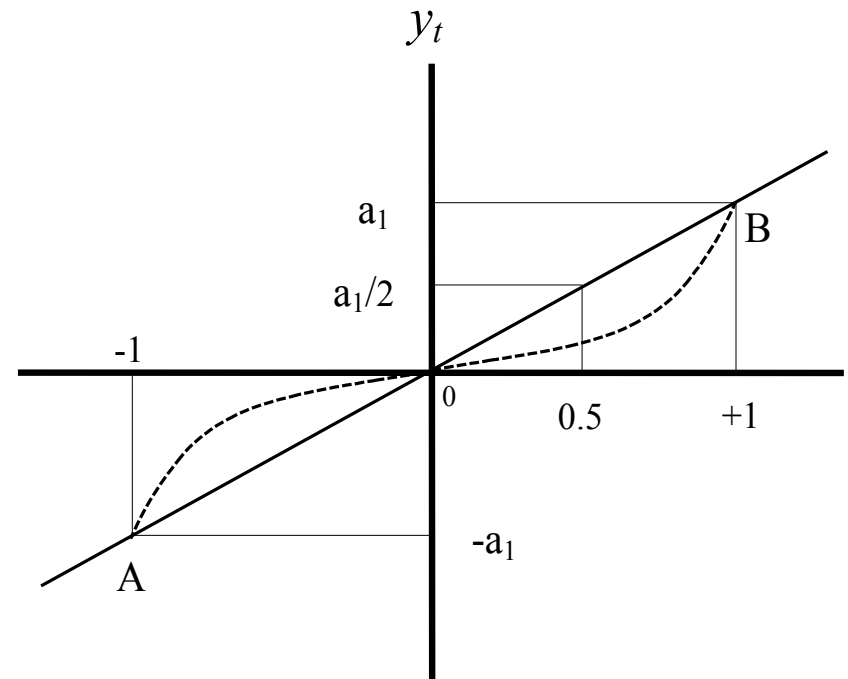
Symmetric vs Asymmetric Adjustment

y_t

$y_t = \rho_1 y_{t-1}$

-1

$\left.\vphantom{\rule{0pt}{1em}}\right\} \rho_1$

y_{t-1}

$-\rho_1 \left\{\vphantom{\rule{0pt}{1em}}\right.$

+1

ρ_2

Figure 7.1: Two Nonlinear Adjustment Paths

Panel a

Panel b

Autoregressive Moving Average (ARMA) Models

The standard ARMA (p, q) model has the form:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i}$$

- ARMA (p, q) models have popularized by Box and Jenkins
- The main econometric problem is to determine the lag lengths p and q and then estimate the parameters α_i and β_i. If all $\beta_i = 0$, the ARMA model is a pure autoregressive (AR) model of order p.
- The second econometric problem is to determine the degree of differencing that is appropriate to render $\{y_t\}$ stationary.
- The key point to note is that the ARMA model is linear; all values of y_{t-i} and ε_{t-i} are raised to the power 1 and there are no cross-products of the form of $y_{t-i}\, \varepsilon_{t-j}$ or $y_{t-i} y_{t-j}$

The NLAR(p) Model

- The p-th order nonlinear autoregressive model is:

$$y_t = f(y_{t-1}, y_{t-2}, \ldots, y_{t-p}) + \varepsilon_t$$

- For an NLAR(2), a Taylor series expansion is

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_{12} y_{t-1} y_{t-2} + a_{11} y_{t-1}^2 + a_{22} y_{t-2}^2$$
$$+ a_{112} y_{t-1}^2 y_{t-2} + a_{122} y_{t-1} y_{t-2}^2 + a_{111} y_{t-1}^3 + a_{222} y_{t-2}^3 + \varepsilon_t$$

Generalized Autoregressive (GAR) Models

The general form of a GAR model is:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \, y_{t-i} + \sum_{i=1}^{q} \sum_{j=1}^{r} \sum_{k=1}^{s} \sum_{l=1}^{u} \beta_{ijkl} \, y_{t-i}^{k} \, y_{t-j}^{l} + \varepsilon_t$$

where: p, q, r, s, and u are integers that are greater or equal to 1.

- GAR models extend AR models by adding various powers of lagged values and cross-products of y_{t-i}. Since GAR models are linear in their parameters, they can be estimated using OLS.

- You can use traditional t-tests and F-tests to pare down the number of parameters estimated. However, this can be tricky since the regressors are likely to be highly correlated. As such, the usual practice is to pare down the equation using the AIC or SBC.

Bilinear Autoregressive (BL) Models

- The general form of the bilinear model BL (p, q, r, s) is:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} + \sum_{i=1}^{r}\sum_{j=1}^{s} c_{ij} y_{t-i} \varepsilon_{t-j}$$

- Bilinear models are a natural extension of ARMA models in that they add the crossproducts of y_{t-i} and ε_{t-j} to account for non-linearity. If all values of c_{ij} equal zero, the bilinear model reduces to the linear ARMA model. Priestley (1980) argues that bilinear models can approximate any reasonable non-linear relationship.

- The bilinear model can be viewed as having stochastic parameter variation
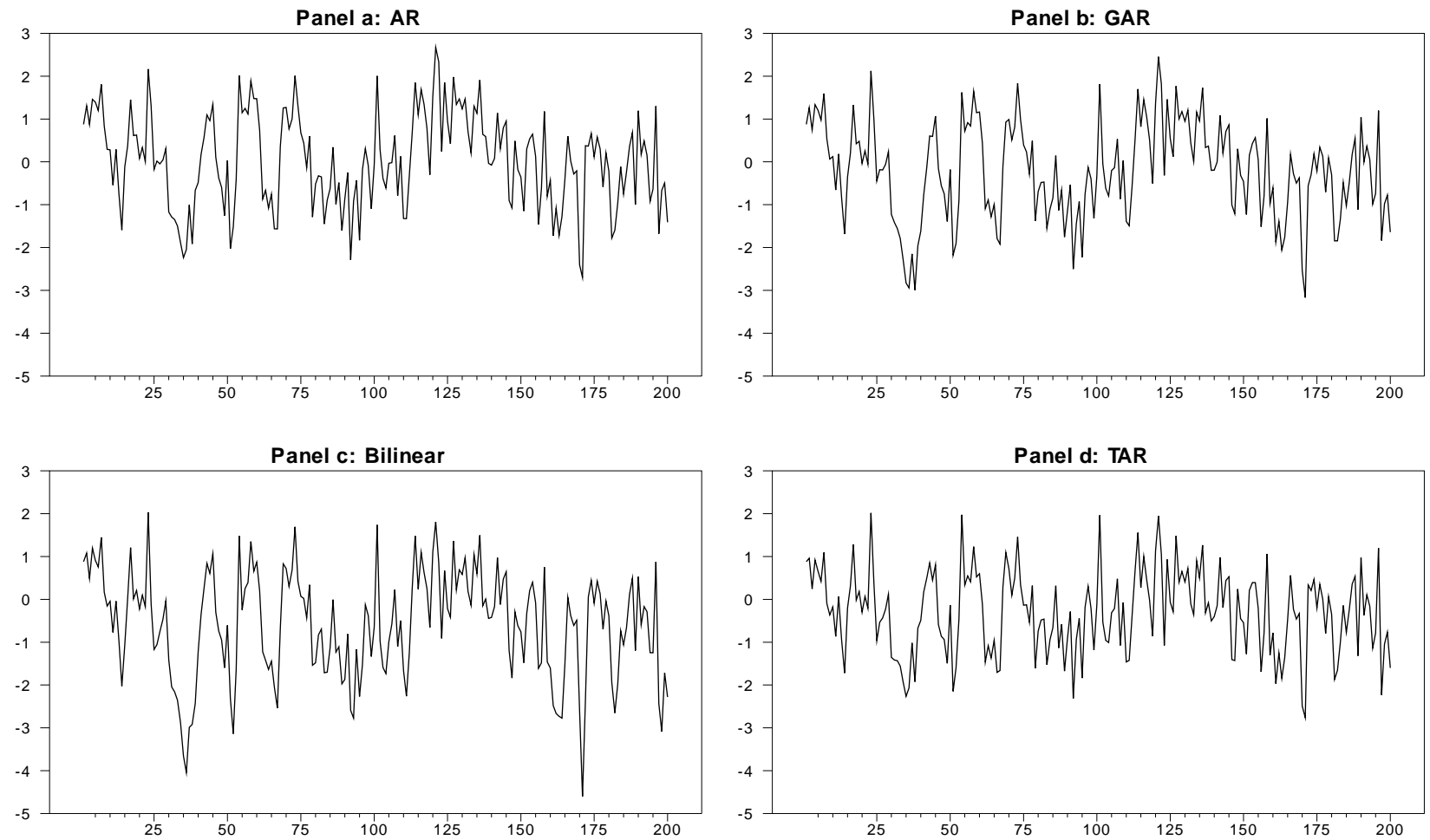 - This is equivalent to a model with ARCH effects

Figure 7.2: Comparison of Linear and Nonlinear Processes

Rothman's unemployment estimates (1998)

AR $\qquad u_t = 1.563u_{t-1} - 0.670u_{t-2} + \varepsilon_t$
$\qquad\qquad$ (22.46) $\qquad$ (−10.06)

GAR $\qquad u_t = 1.500u_{t-1} - 0.553u_{t-2} - 0.745\,(u_{t-2})^3 + \varepsilon_t \qquad$ variance ratio = 0.965
$\qquad\qquad$ (23.60) $\qquad$ (−6.72) $\qquad$ (−2.33)

BL $\qquad u_t = 1.910u_{t-1} - 0.690u_{t-2} - 0.585u_{t-1}\varepsilon_{t-3} + \varepsilon_t \qquad$ variance ratio = 0.936
$\qquad\qquad$ (24.11) $\qquad$ (−10.55) $\qquad$ (−2.08)

where $\quad u_t$ = the detrended log of the unemployment rate over the 1948Q1 to 1979Q4 period

The AIC was used to select the most appropriate values of *p* and *q*

Rothman II

- It is instructive to write the estimated GAR model as

$$u_t = 1.500u_{t-1} - [0.553 + 0.745(u_{t-2})^2] \, u_{t-2} + \varepsilon_t$$

As such, large deviations are less persistent. For the bilinear model:

$$u_t = 1.910u_{t-1} - 0.690u_{t-2} - 0.585u_{t-1}\varepsilon_{t-3} + \varepsilon_t$$

Rothman indicates that u_{t-1} and ε_{t-3} are positively correlated. Since the coefficient on $u_{t-1}\varepsilon_{t-3}$ is negative, large shocks to the unemployment rate imply a faster speed of adjustment than small shocks. As u_{t-1} and ε_{t-3} tend to move together, the larger $u_{t-1}\varepsilon_{t-3}$, the smaller is the degree of persistence.

Exponential Autoregressive (EAR) Models

- EAR models were examined extensively by Ozaki and Oda (1978), Haggan and Ozaki (1981) and Lawrance and Lewis (1980). A standard form of the EAR model is:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \theta_i \, y_{t-i} + \varepsilon_t$$

$$\theta_i = \alpha_i + \beta_i \exp\left(-\gamma y_{t-1}^2\right) \qquad \text{and } \gamma > 0 \text{ is the smoothness parameter.}$$

- In the limit as γ approaches zero or infinity, the EAR model becomes an AR(p) model since each θ_i is constant. Otherwise, the EAR model displays non-linear behavior. For example, equation can capture a situation in which the periods surrounding the turning points of a series (*i.e.*, periods in which y_{t-1}^2 will be extreme) have different degrees of autoregressive decay than other periods.
- Note that adjustment is symmetric but nonlinear
- This is a special case of the ESTAR model to be considered later

The ACF Can be Misleading in a Nonlinear Model

Correlation is only a measure of linear association. Consider:

$$y_t = \beta y_{t-1} x_{t-1} + \mu_t$$

where: y_t is observable but μ_t and x_t are both white noise.
Here, all ρ_k (for $k > 0$) are zero.

$$E[y_t \, y_{t-k}] = \beta^2 E[(y_{t-1} x_{t-1} + \mu_t)(y_{t-k} x_{t-k} + \mu_{t-k})]$$
$$= \beta^2 * 0$$

Also, all cross correlations are zero. Consider:

$$E[y_t \, x_{t-k}] = \beta E[(y_{t-1} x_{t-1} + \mu_t) x_{t-k}]$$
$$= \beta * 0 \text{ for } k \neq 1$$
$$= \beta \text{Var}(x) E y_{t-1} = 0$$

However, the optimal non-linear one-step ahead forecast is: $\beta y_t x_t$.

- Also, data generated by an explosive process AR(1) process will have an ACF like that from a stationary AR(1) process.

Some Tests for Nonlinearity

- McLeod–Li (1983) test: Since we are interested in nonlinear relationships in the data, a useful diagnostic tool is to examine the ACF of the squares or cubed values of a series.

  – Let ρ_i denote the sample correlation coefficient between squared residuals and use the Ljung–Box statistic to determine whether the squared residuals exhibit serial correlation.

$$Q = T(T+2) \sum_{i=1}^{n} \rho_i^2 /(T-i)$$

$$\hat{e}_t^2 = \alpha_0 + \alpha_1 \hat{e}_{t-1}^2 + ... + \alpha_n \hat{e}_{t-n}^2 + v_t$$

Regression Error Specification Test (RESET)

- **STEP 1**: Estimate the best-fitting linear model. Let $\{e_t\}$ be the residuals from the model

- **STEP 2**: Select a value of H (usually 3 or 4) and estimate the regression equation:

$$e_t = \delta z_t + \sum_{h=2}^{H} \alpha_h \hat{y}_t^h$$

where z_t is the vector that contains the variables included in the model estimated in Step 1.

Hence, you can reject linearity if the sample value of the F-statistic for the null hypothesis $\alpha_2 = \ = \alpha_H = 0$ exceeds the critical value from a standard F-table.

The idea is that this regression should have little explanatory power if the model is truly linear.

Specific Testing for Nonlinearity

- Lagrange Multiplier Tests
  - You need not estimate the nonlinear model
  - They have a specific alternative hypothesis
  - Unfortunately, they detect many types of nonlinearity
- Methodology—H_0: The model has a particular linear form against a specific alternative.
  - **Step 1**. Estimate the linear portion of the model to get the residuals e_t (i.e., estimate the model under H_0)
  - **Step 2**. Regress e_t on $\partial f(\ )/\partial \beta$ evaluated at the constrained values of β.
  - **Step 3**. From the regression in Step 2, it can be shown that: $TR^2 \sim \chi^2$ with degrees of freedom equal to the number of restrictions. Thus, if the calculated value of TR^2 exceeds that in a χ^2 table, reject H_0.
    - With a small sample, it is standard to use an F-test.

Example 1

- $y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-1} y_{t-2} + \varepsilon_t$
  - $H_0: a_2 = 0$
- In this case you can estimate the nonlinear model and perform a t-test on a2. However to illustrate the procedure:
- **Step 1**: Estimate the model under H_0 to get the estimated residuals; i.e., estimate
  - $y_t = a_0 + a_1 y_{t-1} + e_t$
- **Step** 2: The partial derivatives of y_t w.r.t. parameters are 1, y_{t-1} and $y_{t-1} y_{t-2}$. Hence, regress the residuals on a constant, y_{t-1} and $y_{t-1} y_{t-2}$
- **Step 3**: Find TR^2. This is χ^2 with 1 degree of freedom

Example 2: Bilinear Model

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + y_{t-2}\varepsilon_{t-1} + \varepsilon_t$$

$$H_o: \gamma = 0$$

- Regress y_t on a constant y_{t-1} and y_{t-2} to obtain e_t
- Regress e_t on a constant, y_{t-1} , $y_{t-2,}$ and $y_{t-2}\varepsilon_{t-1}$

THRESHOLD AUTOREGRESSIVE MODELS

- As in the equation for the spread, if we include a disturbance term, the basic TAR model is

$$y_t = \begin{cases} a_1 y_{t-1} + \varepsilon_{1t} \text{ if } y_{t-1} > 0 \\ a_2 y_{t-1} + \varepsilon_{2t} \text{ if } y_{t-1} \leq 0 \end{cases}$$

- If we assume that the variances of the two error terms are equal [*i.e.*, $\text{var}(\varepsilon_{1t}) = \text{var}(\varepsilon_{2t})$]

$$y_t = a_1 I_t y_{t-1} + a_2 (1 - I_t) y_{t-1} + \varepsilon_t$$

- *where $I_t = 1$ if $y_{t-1} > 0$ and $I_t = 0$ if $y_{t-1} \leq 0$.*
- The indicator can also be set using Δy_t

The Standard TAR Model

- Consider

$$y_t = I_t \left[\alpha_{10} + \sum_{i=1}^{p} \alpha_{1i} y_{t-i} \right] + (1 - I_t) \left[\alpha_{20} + \sum_{i=1}^{r} \alpha_{2i} y_{t-i} \right] + \varepsilon_t$$

- *where* $\quad I_t = 1$ if $y_{t-1} > \tau$ and $I_t = 0$ if $y_{t-1} \leq \tau$.

The M-TAR Model

- The momentum threshold autoregressive (M-TAR) model used by Enders and Granger (1998) allows the regime to change according to the *first-difference* of $\{y_{t-1}\}$. Hence, equation is replaced with:

$$I_t = \begin{cases} 1 & if \quad \Delta y_{t-1} \geq \tau \\ 0 & if \quad \Delta y_{t-1} < \tau. \end{cases}$$

- It is argued that the M-TAR model is useful for capturing situations in which the degree of autoregressive decay depends on the direction of change in $\{y_t\}$.

- Enders and Granger (1998) and Enders and Siklos (2001) show that interest rate adjustments to the term-structure relationship display M-TAR behavior. It is important to note that for the TAR and M-TAR models, if all $\alpha_{1i} = \alpha_{2i}$ the TAR and M-TAR models are equivalent to an AR(p) model.

- See TAR_figure.prg

Extensions

- Selecting the Delay Parameter
- Multiple Regimes
  - **band-TAR**

$$s_t = +a_1(s_{t-1} - \mu) + \varepsilon_t \qquad \text{when } s_{t-1} > +c$$

$$s_t = s_{t-1} + \varepsilon_t \qquad \text{when } -c < s_{t-1} \leq +c$$

$$s_t = +a_2(s_{t-1} - \mu) + \varepsilon_t \qquad \text{when } s_{t-1} \leq -c$$

Estimating τ

- If τ is known, the estimation of the TAR and M-TAR models is straightforward. Simply form the variables $y = I_t y_{t-i}$ and $y = (1 - I_t) y_{t-i}$ and estimate equation using OLS. The lag length p can be determined as in an AR model.
- When τ is unknown, Chan (1993) shows how to obtain a super-consistent estimate of the threshold parameter. For a TAR model, the procedure is to order the observations from smallest to largest such that:

$$y^1 \leq y^2 \leq \ldots \leq y^T$$

For each value of y^i, let $\tau = y^i$ and set the Heaviside indicator accordingly.

Estimate TAR model--the regression equation with the smallest residual sum of squares contains the consistent estimate of the threshold.
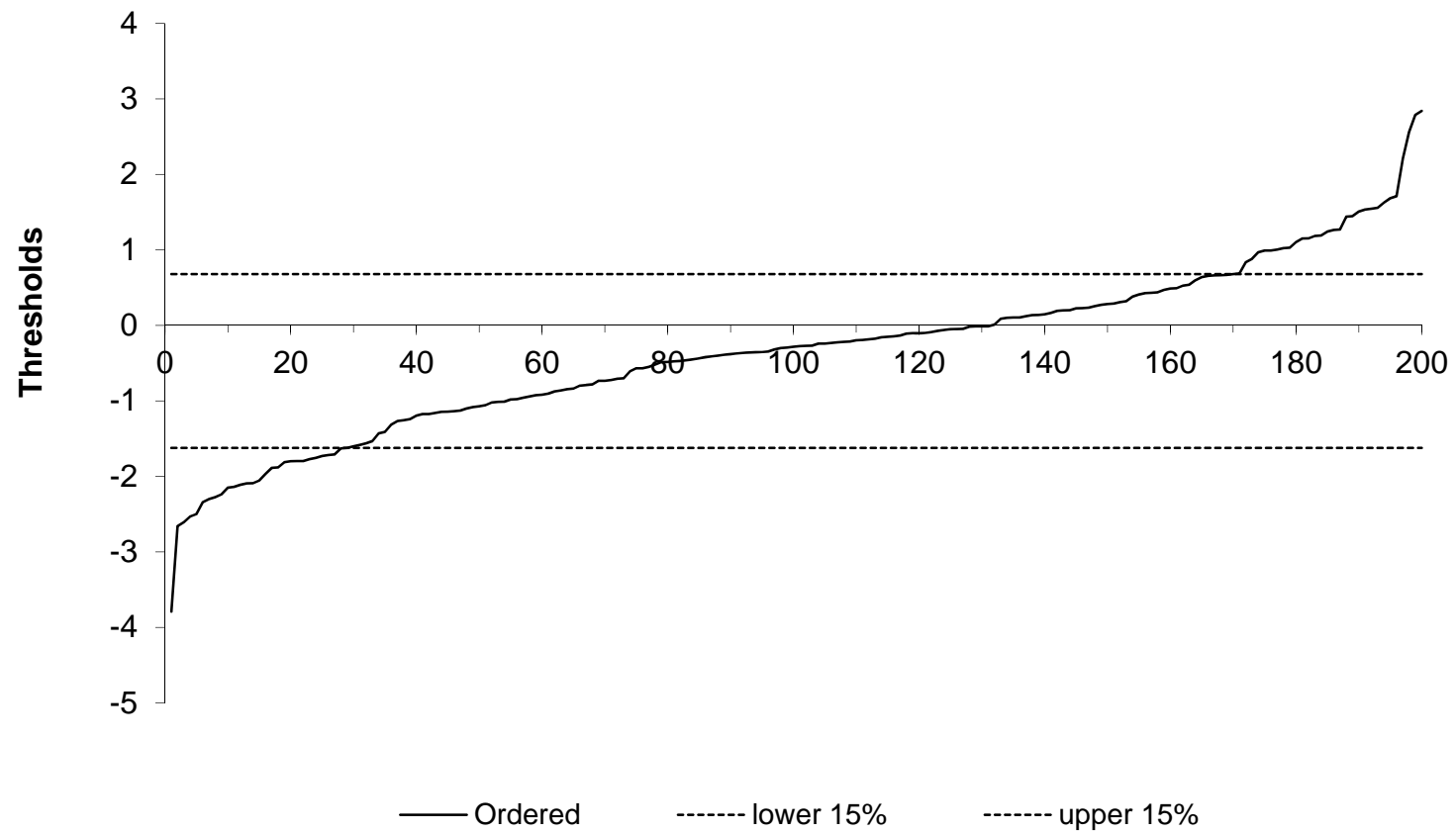
In practice, the highest and lowest 10% of the $\{y\}$ values are excluded from the grid search to ensure an adequate number of observations on each side of the threshold. For the M-TAR model, is replaced by the ordered first-differences of the observations.

Figure 7.3: Estimation of the Threshold

TAR Series ——•—— upper 15% ------- lower 15% -------
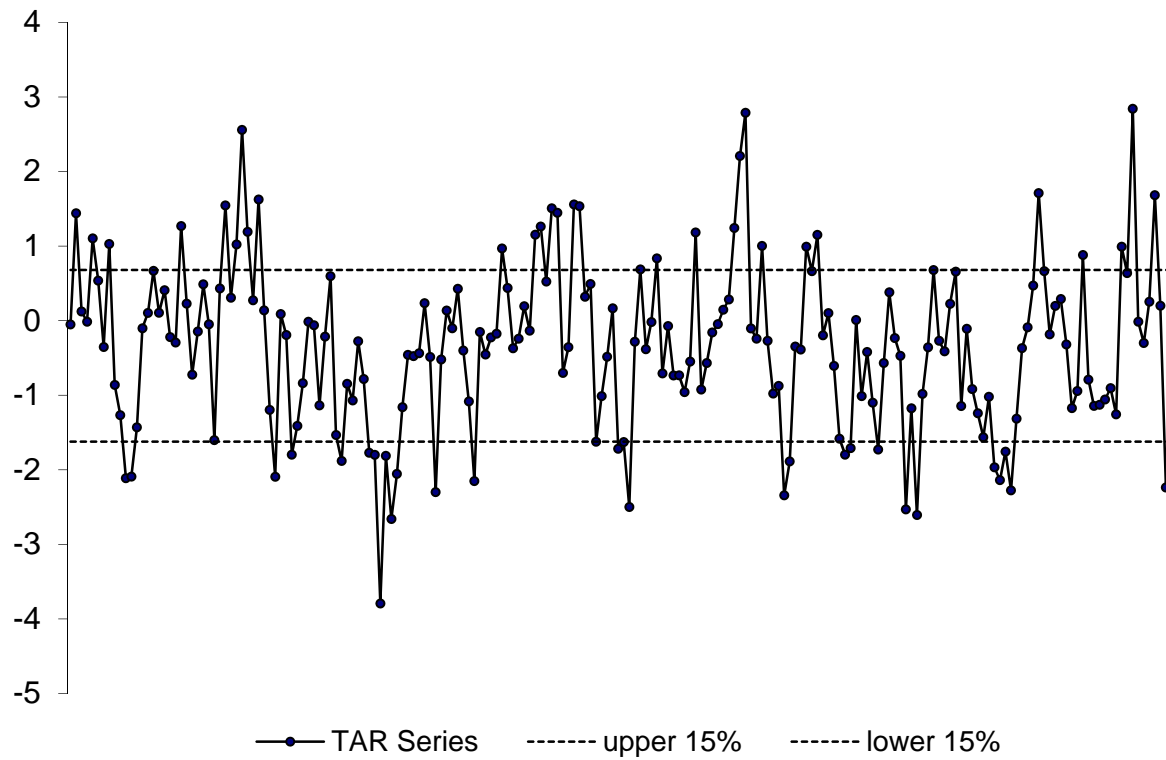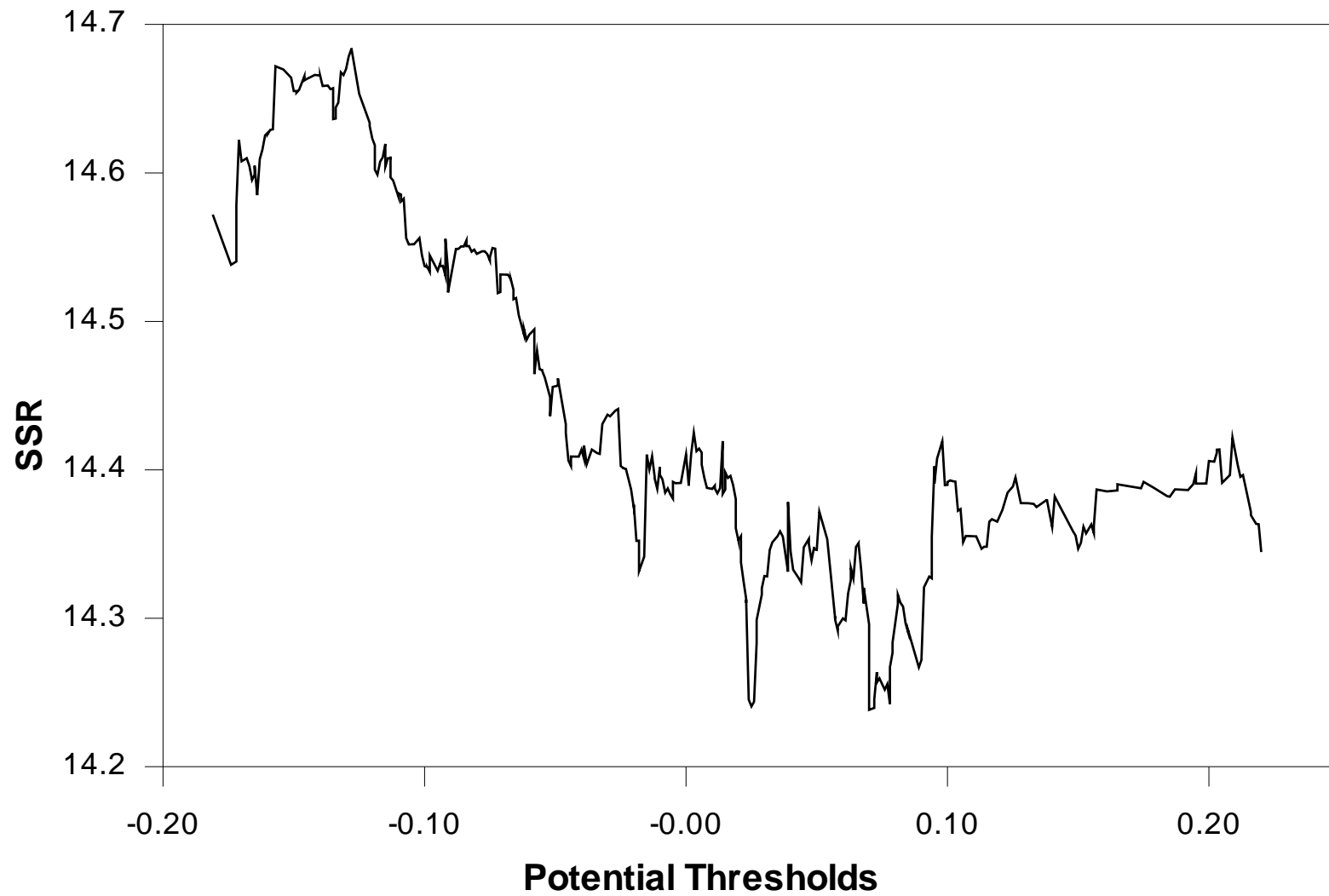
Figure 7.4: Ordered Threshold Values

Figure 7.7: SSR and the Potential Thresholds

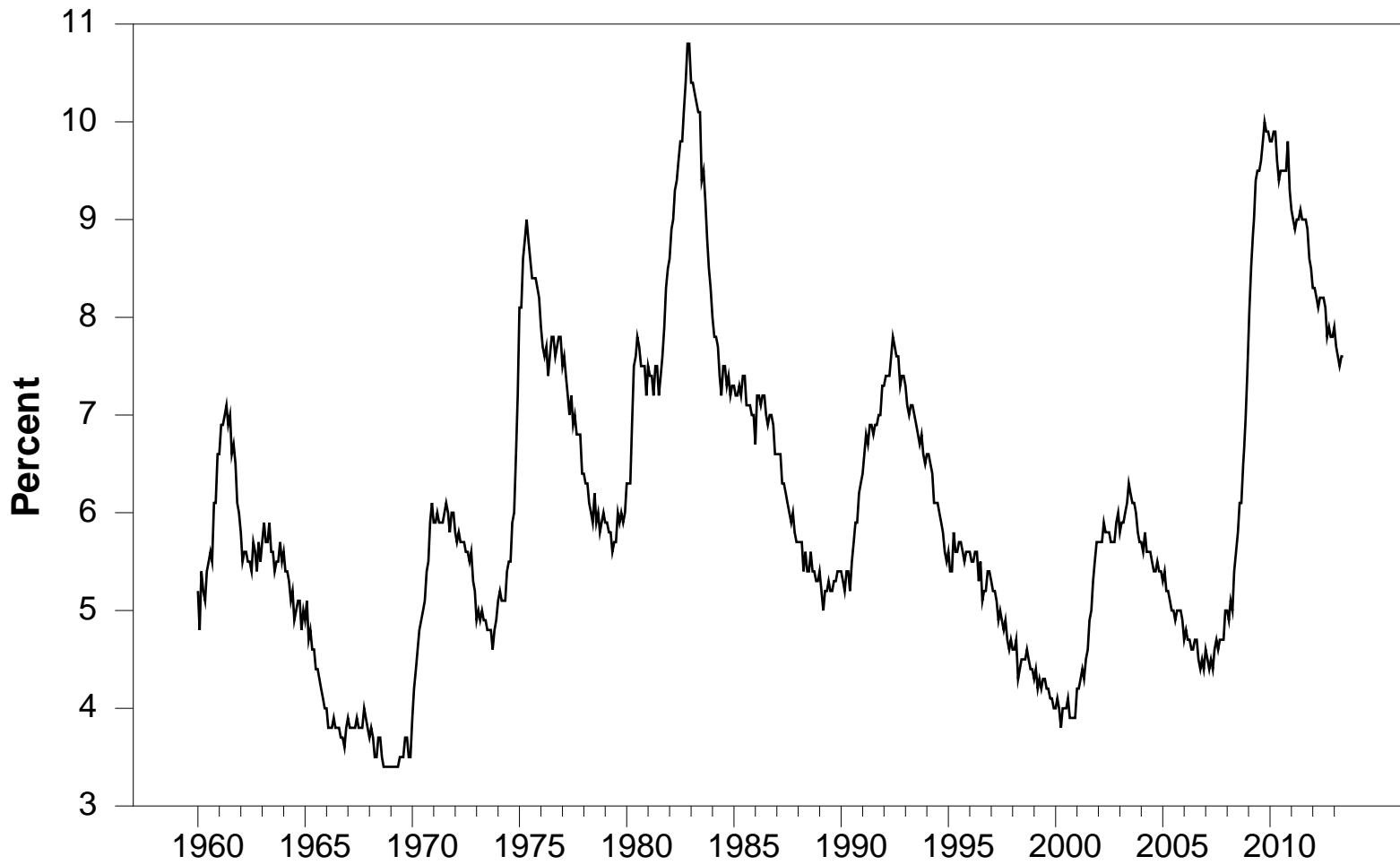Figure 7.6 The U.S. Unemployment Rate

Unidentified Nuisance Parameters

- **Example 1**: $y_t = \alpha_0 + \alpha_1 x_t^{\alpha_2} + \varepsilon_t$

  – Estimate by NLLS. Under the null $\alpha_2 = 0$, the model becomes

  $$y_t = \alpha_0 + \alpha_1 + \varepsilon_t$$

- **Example 2.** $y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 D_t + \varepsilon_t$

  – $D_t = 1$ if $t \geq t^*$ and $D_t = 0$ otherwise. If the break date t^* is unknown, t^* is an unidentified nuisance parameter.

- **Example 3:** $y_t = \alpha_0 + \alpha_1/[1 + \exp(-\gamma y_{t-1})] + \varepsilon_t$.

  – if γ is unknown, a test for linearity implies $\gamma = 0$ so that $y_t = \alpha_0 + \alpha_1/2 + \varepsilon_t$ (since $\exp(0) = 1$).

  – Similarly if $\alpha_1 = 0$, the model becomes $y_t = \alpha_0 + \varepsilon_t$ so that γ is not identified in that its value is irrelevant.

- In a 2-parameter model the log likelihood function can be written solely as a function of α_1 and α_2:
  - $\mathcal{L}(\alpha_1, \alpha_2)$

  Call $\mathcal{L}_a(\alpha_1, \alpha_2)$ this maximized value under the alternative.
- Call $\mathcal{L}_n(\alpha_1{}^*, \alpha_2)$ the restricted value under the null $\alpha_1 = \alpha_1{}^*$
- Let $r = 2[\mathcal{L}_a(\alpha_1, \alpha_2) - \mathcal{L}_n(\alpha_1{}^*, \alpha_2)]$ whch should equal zero.
- If α_2 is not identified under the null hypothesis
  - $r = 2[\mathcal{L}_a(\alpha_1, \alpha_2) - \mathcal{L}_n(\alpha_1{}^*)]$

  which depends on α_2

r does not have a standard χ^2 distribution

Inference

- Inference on the coefficients in a threshold model is not straightforward since it was necessary to search for τ. Under the null of linearity γ is not identified.

- The *t*-statistics yield only an approximation of the actual significance levels of the coefficients. The problem is that the coefficients on the various Δu_{t-i} are multiplied by I_t or $(1-I_t)$ and that these values are dependent on the estimated value of τ.

- The percentile and bootstrap *t* methods can be used

Hansen's (1997) **supremum test**.

- You cannot perform a traditional F-test.
- To use Hansen's (1997) bootstrapping method, you need to draw T normally distributed random numbers with a mean of zero and a variance of unity; let e_t denote this set of random numbers. You treat e_t as the dependent variable. Regress e_t on the actual values of y_{t-i} to obtain an estimate of SSR_r called SSR_r^*.
- For each potential value of τ regress e_t on $I_t y_{t-1}$ and $(1-I_t)y_{t-1}$ [i.e., estimate a regression in the form $e_t = \alpha I_t y_{t-1} + \beta(1-I_t)y_{t-1}$] and use the regression providing the best fit. Call the sum of squared residuals from this regression SSR_u^*. Use these two sums of squares to form

$$F^* = \frac{(SSR_r^* - SSR_u^*)/n}{(SSR_u^*/(T-2n))}$$

- Repeat this process several thousand times to obtain the distribution of F^*.

Threshold Regression Models

- $y_t = a_0 + (a_1 + b_1 I_t)x_t + \varepsilon_t$

  where $\quad I_t = 1$ if $y_{t-d} > \tau$ and $I_t = 0$ otherwise.

- **Pretesting for a TAR Model**

$$F = \frac{(SSR_r - SSR_u)/n}{(SSR_u/(T - 2n))}$$

  However, the F-statistic needs to be bootstrapped. (see Hansen above)

TAR Models and Endogenous Breaks

- The threshold model is equivalent to a model with a structural break. The only difference is that in a model with structural breaks, *time* is the threshold variable.
- Carrasco (2002) shows that the usual tests for structural breaks (i.e., those using dummy variables) have little power if the data are actually generated by a threshold process
  - However, a test for a threshold process using y_{t-d} as the threshold variable has power to detect both threshold behavior *and* structural change. Even if there is a single structural break at time period t, using y_{t-d} as the threshold variable will mimic this type of behavior.
  - As such, she recommends using the threshold model as a general test for parameter instability.

Asymmetric Monetary Policy

- Consider the linear Taylor Rule

$$i_t = -0.269 + 0.464\pi_t + 0.345y_t + 0.810i_{t-1}$$
$$(-1.47) \quad (6.05) \qquad (5.16) \qquad (21.83)$$

$$\text{AIC} = -27.72 \text{ and SBC} = -16.85$$

The TAR Taylor Rule

- Since we do not know the delay factor, we can estimate four threshold regressions with π_{t-1}, π_{t-2}, y_{t-1} and y_{t-2} as the threshold variables

	τ	*SSR*	*AIC*	*BIC*
π_{t-1}	3.527	50.80	−70.55	−46.08
π_{t-2}	3.668	50.42	−71.39	−46.93
y_{t-1}	−1.183	63.97	−44.73	−20.26
y_{t-2}	−1.565	53.51	−64.94	−40.47

$$i_t = 1.421 + 1.051\pi_t + 0.469y_t + 0.374i_{t-1} \text{ when } \pi_{t-2} \geq 3.527$$
$$\quad (3.15) \quad (10.55) \quad\quad (6.22) \quad\quad (5.74)$$

and

$$i_t = -0.456 + 0.232\pi_t + 0.302y_t + 0.959i_{t-1} \text{ when } \pi_{t-2} < 3.527$$
$$\quad (-1.40) \quad (1.88) \quad\quad (3.77) \quad\quad (24.55)$$

Capital Stock Adjustment with Multiple Thresholds

- For our purposes, the key variables in the Boetel, Hoffman and Liu (2007) model are

$$K_t - K_{t-1} = 4569 + 6360\ I_{1t} + 6352\ I_{2t} + 452p_{Ht-1} - 2684p_{Ft-1} + \ldots$$
$$\phantom{K_t - K_{t-1} = }(3.30)\quad(5.59)\quad\quad(5.20)\quad\quad(1.84)\quad\quad(-3.66)$$

- where: K_t is the size of the breeding stock, p_{Ht-1} is a measure of the output price of hogs, and p_{Ft-1} is a measure of the price of feed. The indicators functions are such that $I_{1t} = 1$ if $p_{Ht-1} > \tau_{high} = 1.1185$ and $I_{2t} = -1$ if $p_{Ht-1} < \tau_{low} = 1.1105$. The use of lagged values for the dependent variables is designed to reflect a one period delay between the time of the investment decision and its realization.

  - allowing all variables to have asymmetric effects on $K_t - K_{t-1}$ would entail estimating a large number of parameters with a consequent loss of degrees of freedom.

Capital Stock Adjustment with Multiple Thresholds II

- … the three regimes are distinguished by p_{Ht-1} relative to two threshold values.
- when p_{Ht-1} is between τ_{high} and τ_{low}, I_{1t} and $I_{2t} = 0$ so that the intercept is 4569.
- when $p_{Ht-1} > \tau_{high}$, $I_{1t} = 1$ the intercept is 10929 and when $p_{Ht-1} < \tau_{low}$, $I_{2t} = -1$ the intercept is 8.
- Thus, there is a high-, sluggish- and disinvestment regime whose presence is dependent on the value of p_{Ht-1}.
- It would be a mistake to conclude that the slope coefficient 452 measures the full effect of a price change on net investment. When the value of p_{Ht-1} crosses one of the thresholds, the change in investment is enhanced since the intercept changes along with the price. Also note that price changes within the interval τ_{high} to τ_{low}, will little effect on investment.

Smooth Transition Models

- In some instances, it may not be reasonable to assume that there are 2 pure regimes:
  - Multi-regime TAR model
  - It is possible to assume that the transition is smooth
- Smooth transition autoregressive (STAR) models allow the autoregressive parameters to change slowly:
  - $y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 y_{t-1} f[y_{t-1} - \mu] + \varepsilon_t$

  where: $f[\bullet]$ is a continuous function.

  Typically:

  $f(0) = 1$ and $f(\pm\infty) = 0$ (as in a density function).

  when $y_{t-1} = \mu$, autoregressive decay is $\alpha_1 + \beta_1$ and when $|y_{t-1} - \mu|$ is large, decay is α_1.

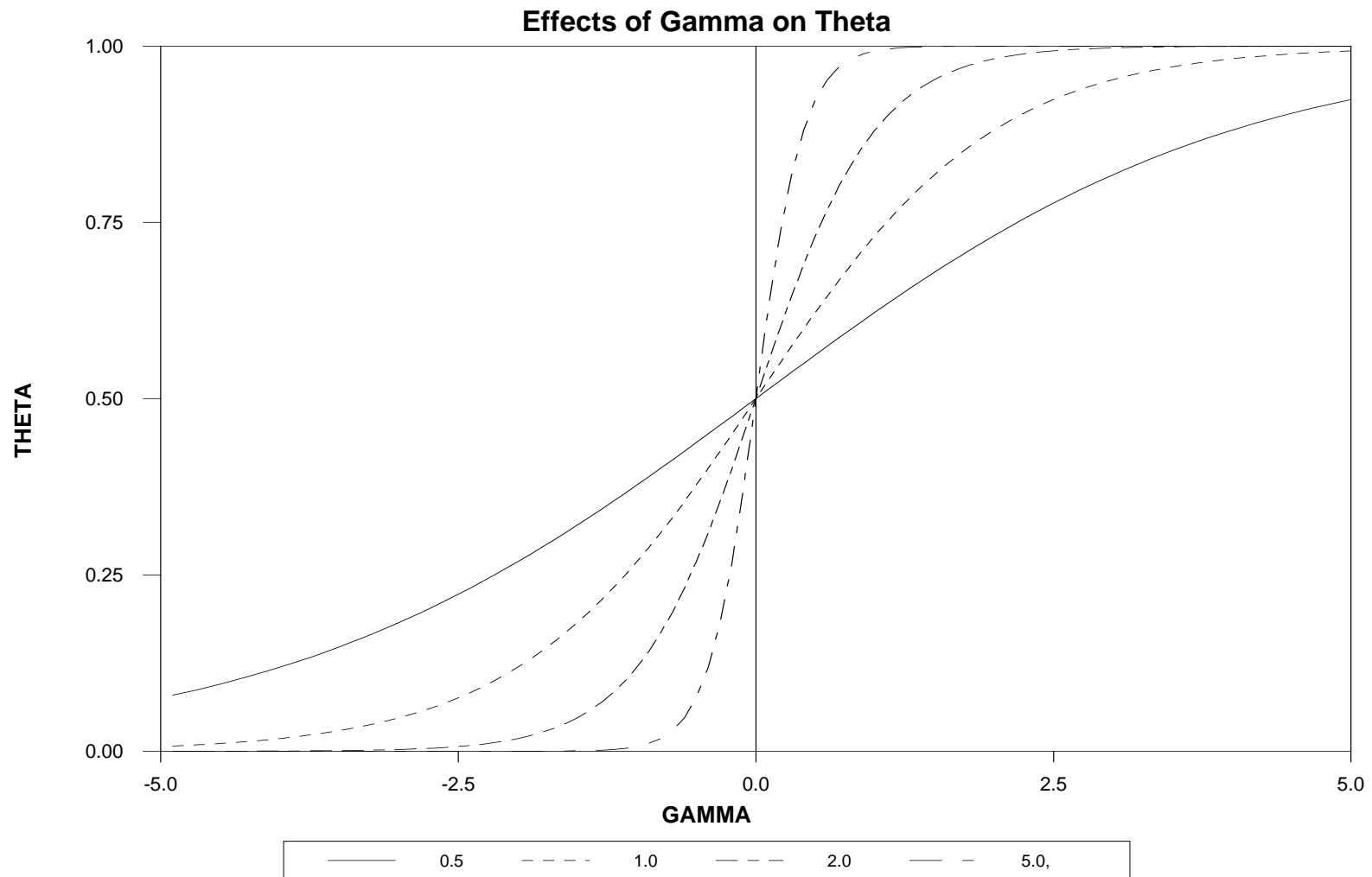# The Logistic Smooth Transition Autoregressive (LSTAR) Model

- The LSTAR model generalizes the standard autoregressive model to allow for a varying degree of autoregressive decay.

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \theta(\beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i}) + \varepsilon_t$$

$$\theta = [1 + \exp(-\gamma y_{t-1} - c)]^{-1} \quad \text{and } \gamma > 0 \text{ is a scale parameter}$$

- In the limit as $\gamma \to 0$ or ∞, the LSTAR model becomes an $AR(p)$ model since each value of θ is constant.

- For intermediate values of γ, the degree of autoregressive decay depends on the value of y_{t-1}. As $y_{t-1} \to -\infty$, $\theta \to 0$ so that the behavior of y_t is given by $\alpha_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \varepsilon_t$. As $y_{t-1} \to +\infty$, $\theta \to 1$ so that the behavior of y_t is given by $(\alpha_0 + \beta_0) + (\alpha_1 + \beta_1) y_{t-1} + \ldots + (\alpha_p + \beta_p) y_{t-p} + \varepsilon_t$.

The LSTAR 'AR' Coefficient θ

Effects of Gamma on Theta

Legend: 0.5 — — — 1.0 — — 2.0 — — 5.0,

Pretesting for an LSTAR Model

For the LSTAR model:

$$\theta = [1 + \exp(-\gamma(y_{t-d} - c))]^{-1} \equiv [1 + \exp(-h_{t-d})]^{-1}$$

Use a third-order Taylor series approximation of θ with respect to h_{t-d} evaluated $h_{t-d} = 0$. Of course, this is identical to evaluating the expansion at $\gamma = 0$.

$\partial\theta / \partial h_{t-d}$	$\exp(-h_{t-d})/[1 + \exp(-h_{t-d})]^2$	$1/4$
2nd deriv.	$\exp(-h_{t-d})[1 - \exp(-h_{t-d})]/[1 + \exp(-h_{t-d})]^3$	0
3rd deriv	$\exp(-h_{t-d})[1+\exp(-2h_{t-d}) - 4\exp(-h_{t-d})]/[1 + \exp(-h_{t-d})]^4$	$-1/8$

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + (\beta_0 + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p})(\pi_1 h_{t-d} + \pi_3 (h_{t-d})^3) + \varepsilon_t$$

Because h_{t-d} depends only on the value of y_{t-d}, we can write the model in the more compact form:

can test for the presence of LSTAR behavior by estimating an auxiliary regression:

$$e_t = a_0 + a_1 y_{t-1} + \ldots + a_p y_{t-p} + a_{11} y_{t-1} y_{t-d} + \ldots + a_{1p} y_{t-p} y_{t-d} + a_{21} y_{t-1} + \ldots + a_{2p} y_{t-1}$$
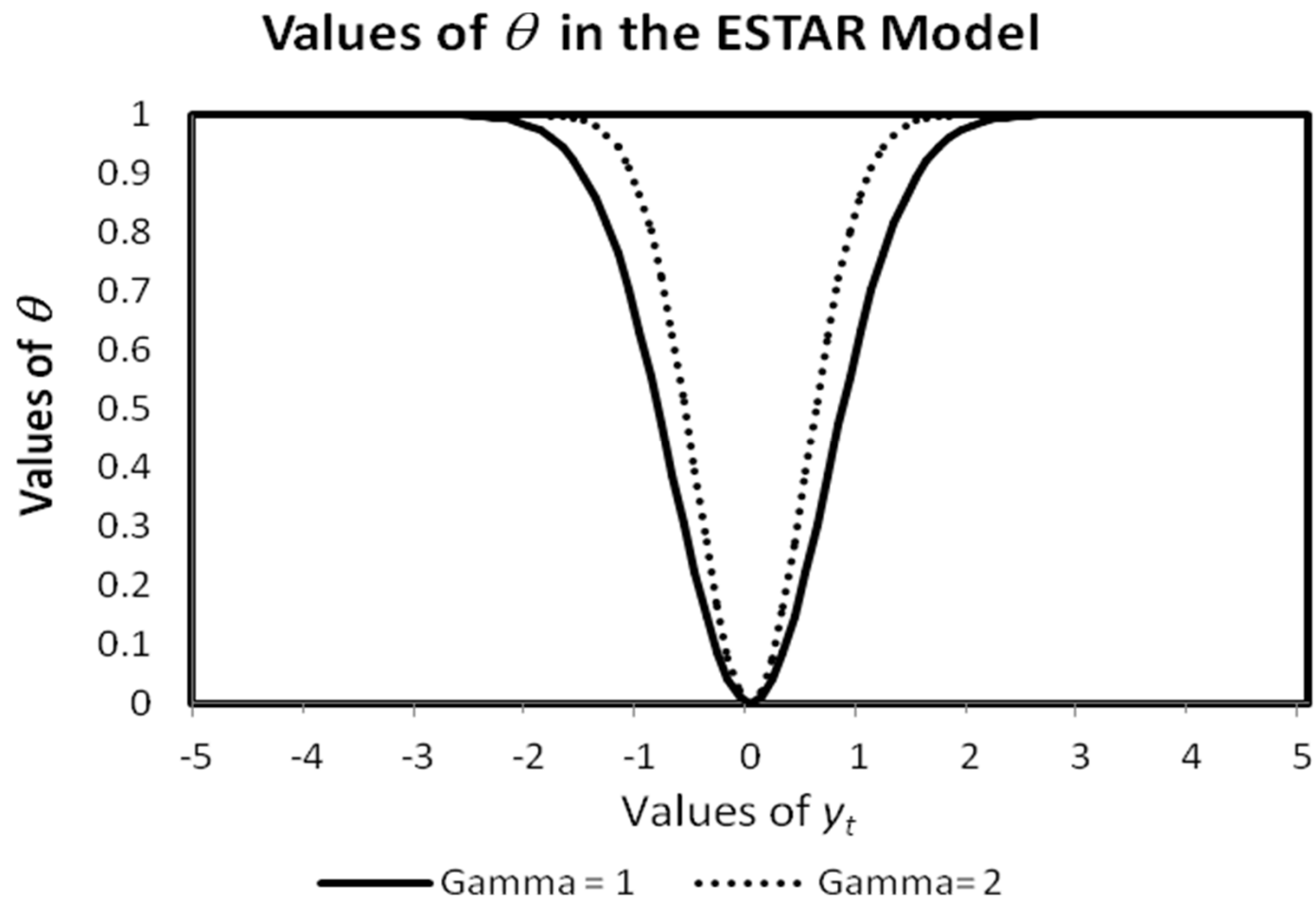$$+ a_{31} y_{t-1} + \ldots + a_{3p} y_{t-p} + \varepsilon_{t.} \quad (7.21)$$

The ESTAR Model

- The exponential form of the model (ESTAR) uses (7.19), but replaces (7.20) with

$$\theta = 1 - \exp\left[-\gamma(y_{t-1} - c)^2\right] \quad \gamma > 0.$$

Note that for an ESTAR model Timo's test has a quadratic term but not a cubic term

Values of θ in the ESTAR Model

Gamma = 1 ⋯⋯ Gamma = 2

Pretesting for an ESTAR Model

Let θ be: $\theta = 1 - \exp(-h^2_{t\text{-}d})$ so that $h_{t\text{-}d} = \gamma^{1/2}(y_{t-d} - c)$

Now, the partial derivatives are given by:

	Equals	Evaluated at $h_{t-d} = 0$
$\partial\theta / \partial h_{t-d}$	$2h_{t-d}\exp(-h^2_{t-d})$	0
$\partial^2\theta / \partial h^2_{t-d}$	$2\exp(-h^2_{t-d}) - 4h^2_{t-d}\exp(-h^2_{t-d})$	2
$\partial^3\theta / \partial h^3_{t-d}$	$-12h_{t-d}\exp(-h^2_{t-d}) + 8h^3_{t-d}\exp(h^2_{t-d})$	0

Teräsvirta's (1994) Pretest

The key insight in Teräsvirta (1994) is that the auxiliary equation for the ESTAR model is nested within that for an LSTAR model. If the ESTAR is appropriate, it should be possible to exclude all of the terms multiplied by the cubic expression from the Taylor series expansin. Hence, the testing procedure follows these steps:

- **STEP 1**: Estimate the linear portion of the AR(p) model to determine the order p and to obtain the residuals $\{e_t\}$.

- **STEP 2**: Estimate the auxiliary equation (7.21). Test the significance of the entire regression by comparing TR^2 to the critical value of χ^2. If the calculated value of TR^2 exceeds the critical value from a χ^2 table, reject the null hypothesis of linearity and accept the alternative hypothesis of a smooth transition model. (Alternatively, you can perform an F-test).

- **STEP 3**: If you accept the alternative hypothesis (*i.e.*, if the model is nonlinear), test the restriction $a_{31} = a_{32} = = a_{3n} = 0$ using an F-test. If you reject the hypothesis $a_{31} = a_{32} = = a_{3n} = 0$, the model has the LSTAR form. If you accept the restriction, conclude that the model has the ESTAR form.

Estimation Issues

- Many of the numerical methods used to estimate the parameter values have difficulty in simultaneously finding γ and c.
    - It is crucial to provide the numerical routine with very good initial guesses.
    - Estimate γ using a grid search. Fix γ at its smallest possible value and estimate all of the remaining parameters using NLLS. Slightly increase the value of γ and reestimate the model. Continue this process until the plausible values of γ are exhausted. Use the value of γ yielding the best fit.
    - If γ is large and convergence to a solution is a problem, it could be easier to estimate a TAR model instead of the LSTAR model.
    - Terasverta (1994) notes that rescaling the expressions in θ can aid in finding a numerical solution. For example, with an LSTAR model, standardize by dividing $\exp[-\gamma(y_{t-d} - c)]$ by the standard deviation of the $\{y_t\}$ series. For an ESTAR model, standardize by dividing $\exp[-\gamma(y_{t-d} - c)^2]$ by the variance of the $\{y_t\}$ series. In this way, the threshold value c is measured in standardized units so that a reasonable value for the initial guess (e.g., $c = 1$ standard deviation) can be readily made.

Michael, Nobay, and Peel (1997)

- $\Delta y_t = 0.40\Delta y_{t-1} + [1 - \exp(-532.4(y_{t-1} - 0.038)^2)] (-y_{t-1}$
  $+ 0.59\Delta y_{t-2} + 0.57\Delta y_{t-4} - 0.017)$

  The point estimates imply that when the real rate is near 0.038, there is no tendency for mean reversion since $a_1 = 0$. However, when $(y_{t-1} - 0.038)^2$ is very large, the speed of adjustment coefficient is quite rapid. Hence, the adjustment of the real exchange rate is consistent with the presence of transaction costs.

11. UNIT ROOTS AND NONLINEARITY

- $\Delta y_t = I_t \rho_1 (y_{t-1} - \tau) + (1 - I_t) \rho_2 (y_{t-1} - \tau) + \varepsilon_t$ (7.30)

$$I_t = \begin{cases} 1 & if \quad y_{t-1} \geq \tau \\ 0 & if \quad y_{t-1} < \tau \end{cases}$$

STEP 1: If you know the value of τ (for example $\tau = 0$), estimate (7.30). Otherwise, use Chan's method: select the value of τ from the regression containing the smallest value for the sum of squared residuals.

STEP 2: If you are unsure as to the nature of the adjustment process, repeat Step 1 using the M-TAR model.

STEP 3: Calculate the *F*-statistic for the null hypothesis $\rho_1 = \rho_2 = 0$. For the TAR model, compare this sample statistic with the appropriate critical value in Table G.

STEP 4: If the alternative hypothesis is accepted (i.e., if there is an attractor), it is possible to test for symmetric versus asymmetric adjustment since the asymptotic joint distribution of ρ_1 and ρ_2 converges to a multivariate normal.

Old School versus New School

'Old School' forecasting techniques, such as exponential smoothing and the Box-Jenkins methodology, do not attempt to explicitly model or to estimate the breaks in the series.

- Exponential smoothing: place relatively large weights on the most recent values of the series.
- The Box-Jenkins: first-difference or second difference the series in order to control for the lack of mean reversion.
  - Differencing can be chosen by the autocorrelation function or by some type of Dickey-Fuller test.

'New School' forecasters attempt to estimate the number and magnitudes of the breaks. Given that the breaks are well-estimated, it is possible to control for the regime shifts when forecasting.

Figure 1: A Persistent Series with Two Breaks
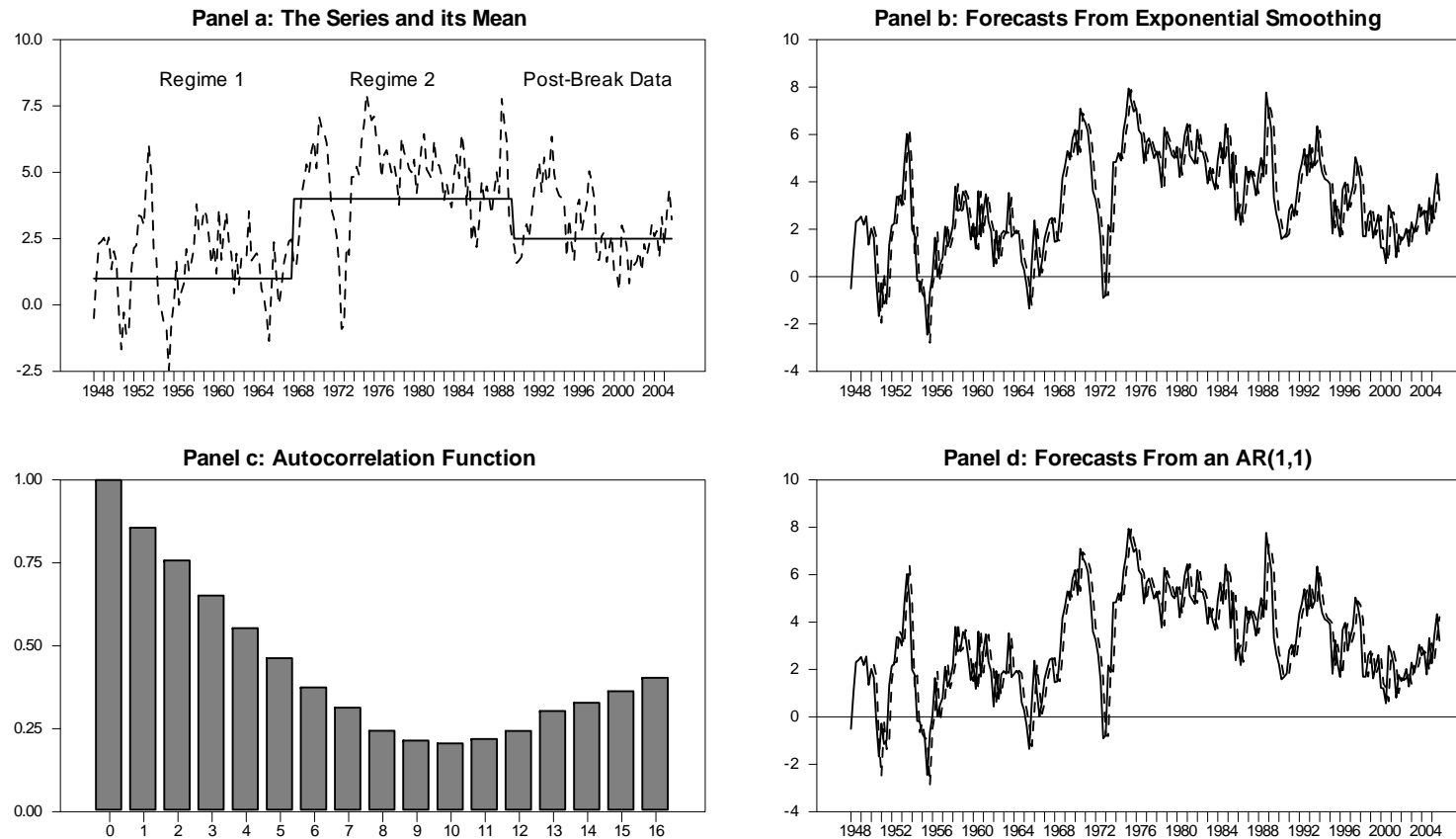
Exponential forecasts place a relatively large weight on the most recent values of the series and quickly captures the mean shift.
Panel d estimates the series in first-differences to remove the effects of the level-shifts from the series.

Endogenous Structural Breaks

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \gamma_0 D_t + \varepsilon_t$$

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + (\gamma_0 + \sum_{i=1}^{p} \gamma_i y_{t-i}) D_t + \varepsilon_t$$

Equation (7.34) is a *partial* break model where the break is assumed to affect only the intercept whereas (7.35) is a *pure* break model in that all parameters are allowed to change. You can use the Andrews and Ploeberger test (1994)

Recall that an endogeneous break model is a threshold model with time as the threshold variable. As such, you can estimate (7.34) or (7.35) by performing a grid search for the best-fitting break date. The test is feasible since the selection of the best fitting regression amounts to a supremum test.

With the sample sizes typically used in applied work, it is standard to use Hansen's (1997) bootstrapping test for a threshold model.

Bai and Peron: Multiple Breaks

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + (\gamma_1 D_{1t} + \gamma_2 D_{2t} + \ldots + \gamma_m D_{mt}) + \varepsilon_t$$

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=1}^{m} D_{jt}\left(\gamma_{0j} + \sum_{i=1}^{p} \gamma_{ij} y_{t-i}\right) + \varepsilon_t$$

Bai and Perron develop a supremum test for the null hypothesis of no structural change ($m = 0$) versus the alternative hypothesis of $m = k$ breaks.

The second method of selecting the number of breaks is to use a sequential test.

Supremum

Estimate models for every possible combination of breaks (given the trimming and minimum break size) and select the best fitting combination of break dates.

The appropriate F-statistic, called the $F(k; q)$ statistic, is nonstandard; the critical values depend on the number of breaks, k, and the number of breaking parameters, q.

If the null hypothesis of no breaks is rejected, they select the actual number of breaks using the SBC. For $q = 1$, 2, and 3, the 95% critical for 1, 2, and 5 breaks are:

q	$k = 1$	$k = 2$	$k = 5$	UDmax
1	9.63	8.78	6.69	10.17
2	12.89	11.60	9.12	13.27
3	15.37	13.84	11.15	16.82

Sequential Method

- Begin with the null hypothesis of no-breaks versus the alternative of a single break. If the null hypothesis of no breaks is rejected, proceed to test the null of a single break versus two breaks, and so forth.

- The method is sequential in that the test for break $\ell + 1$ takes the first ℓ breaks as given. At each stage, the so-called sup $F(\ell+1 \mid \ell)$ statistic is calculated as the maximum F-statistic for the null hypothesis of no additional against the alternative of one additional break. For $q = 1, 2,$ and 3, the 95% critical for $= 0, 1, 2,$ and 5 are:

q	$\ell = 0$	$\ell = 1$	$\ell = 2$	$\ell = 4$
1	9.63	11.14	12.16	13.45
2	12.89	14.50	15.42	16.61
3	15.37	17.15	17.97	19.23

Fourier Breaks (see www.time-series.net)

A simple modification of the standard autoregressive model is to allow the intercept to be a time-dependent function

$$\Delta y_t \;=\; \tilde{\delta}(t) + \sum_{j=1}^{p} \theta_j \Delta y_{t-j} + \rho y_{t-1} + \varepsilon_t, \quad t = 1, \ldots, T,$$

Although (1) is linear in $\{y_t\}$, the specification is reasonably general in that d(t) can be a deterministic polynomial expression in time, a *p-th* order difference equation, a threshold function, or a switching function.

UDmax

it also seems reasonable to test the null of no breaks against the alternative of some breaks. If the largest of the $F(k; q)$ statistics for $k = 1, 2, 3 \ldots$ exceeds the UDmax statistic reported above, you can conclude that there are some breaks and then go on to select the number using the SBC.

The Fourier Approximation

Under very weak conditions, the behaviour of almost any function can be exactly represented by a sufficiently long Fourier series:

$$\delta(t) = \delta_0 + \delta_1 t + \sum_{i=1}^{k}\left[\delta_{si} \sin\frac{2\pi i}{T} \bullet t + \delta_{ci} \cos\frac{2\pi i}{T} \bullet t \right]$$

Note that the linear specification emerges as the special case in which all values of δ_{si} and δ_{ci} are set equal to zero.

Thus, instead of positing a specific nonlinear model, the specification problem becomes one of selection the most appropriate frequencies to include.

Figure 1: Sharp, ESTAR and LSTAR Breaks
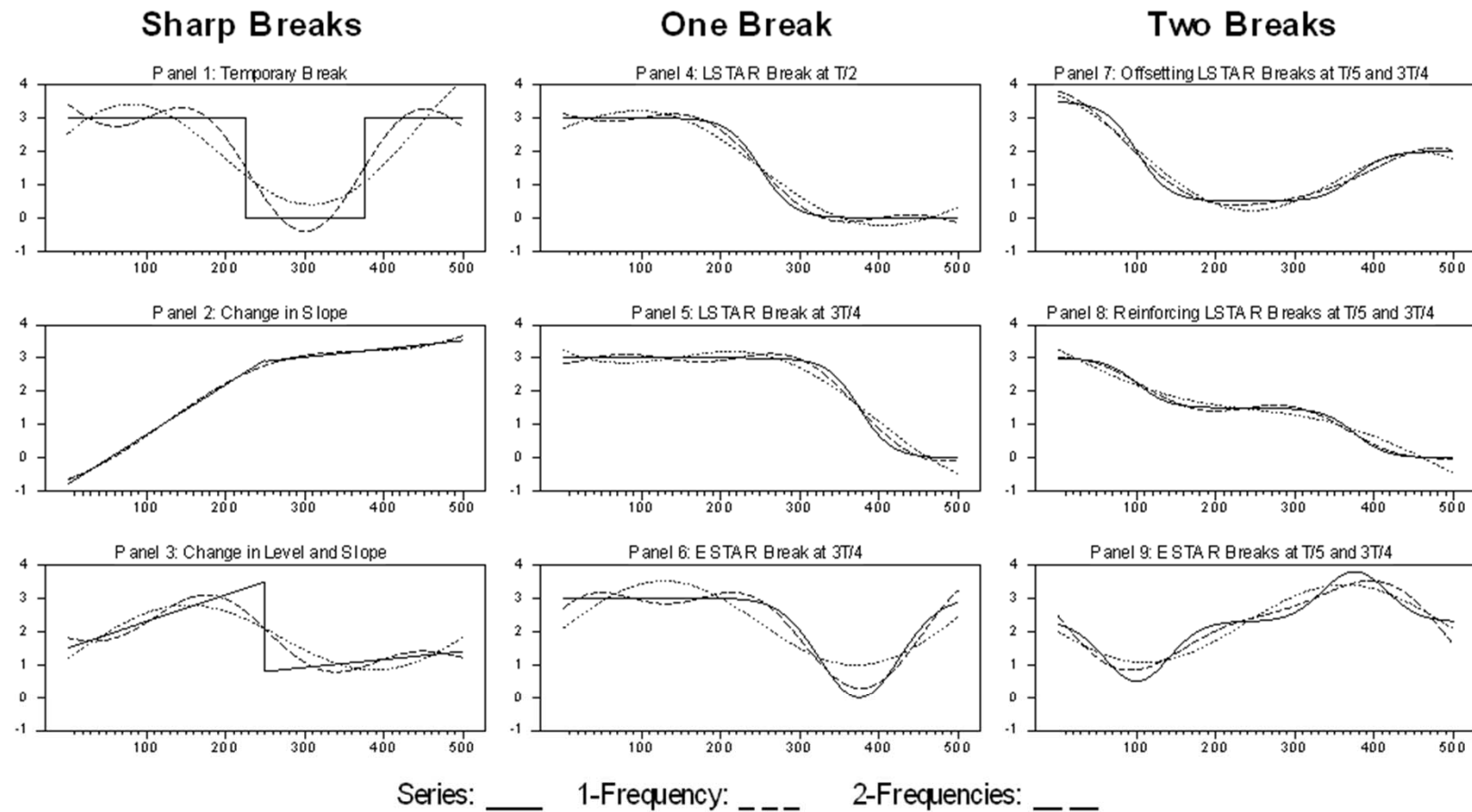
Logistic Breaks

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \; + \alpha_p y_{t-p} + \theta\,[\beta_0 + \beta_1 y_{t-1} + \; + \beta_p y_{t-p}] + \varepsilon_t$$

$$\theta = [1 + \exp(-\gamma(t - t^*))]^{-1}$$

Figure 7.14 A Simulated LSTAR Break

Panel a: Bai-Perron Breaks

Panel b: Logistic Break