Ekonometryczne modele nieliniowe

Wykład 7

Modele łagodnego przejścia,

sieci neuronowe w ekonometrii

Literatura

- Timo Teräsvirta, *Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models*, Journal of the American Statistical Association, Vol. 89, No. 425 (Mar., 1994), pp. 208-218

- Dick van Dijk, Timo Teräsvirta and Philip Hans Franses, *Smooth transition autoregressive models - A survey of recent developments*, Econometric Reviews, 2002, vol. 21, issue 1, pp. 1-47.

Literatura

- Marcelo C. Medeiros & Timo Terasvirta, 2001, *Statistical methods for modelling neural networks*, Textos para discussão 445, Department of Economics PUC-Rio (Brazil).

- Timo Teräsvirta, Dick van Dijk, Marcelo C. Medeiros, *Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination*, International Journal of Forecasting, Volume 21, Issue 4, 2005, pp. 755-774.

Literatura

Książka:

- P.H. Frances, D. van Dijk, *Non-linear time series models in empirical finance*, Cambridge University Press, Cambridge, UK.

Przejście z modelu progowego…

- Model z dwoma reżimami

$$y_t = (\alpha_0 + \alpha_1 x_{1,t} + ... + \alpha_k x_{k,t}) I(z_t \leq \gamma) +$$
$$+ (\beta_0 + \beta_1 x_{1,t} + ... + \beta_k x_{k,t}) I(z_t > \gamma) + \varepsilon_t$$

- …inaczej zapisany

$$y_t = (\alpha_0 + \alpha_1 x_{1,t} + ... + \alpha_k x_{k,t}) +$$
$$+ [(\beta_0 - \alpha_0) + (\beta_1 - \alpha_1) x_{1,t} + ... + (\beta_k - \alpha_k) x_{k,t}] I(z_t > \gamma) + \varepsilon_t$$

$$y_t = \boldsymbol{\alpha}' \mathbf{x}_t + (\boldsymbol{\beta} - \boldsymbol{\alpha})' \mathbf{x}_t I(z_t > \gamma) + \varepsilon_t$$

Model STR

- Smooth Transition (Auto-)Regression

$$y_t = \phi_1' x_t [1 - G(z_t; \gamma, c)] + \phi_2' x_t G(z_t; \gamma, c) + \varepsilon_t \qquad \varepsilon_t \sim i.i.d.(0, \sigma^2)$$

$$y_t = \phi_1' x_t + (\phi_2 - \phi_1)' x_t G(z_t; \gamma, c) + \varepsilon_t$$

Funkcja przejścia

- G – funkcja logistyczna

$$G(z_t;\gamma,c)=\{1+\exp[-\gamma(z_t-c_k)]\}^{-1}$$

- gdy $\gamma=0$, to model liniowy
- gdy $\gamma\rightarrow\infty$, to model progowy

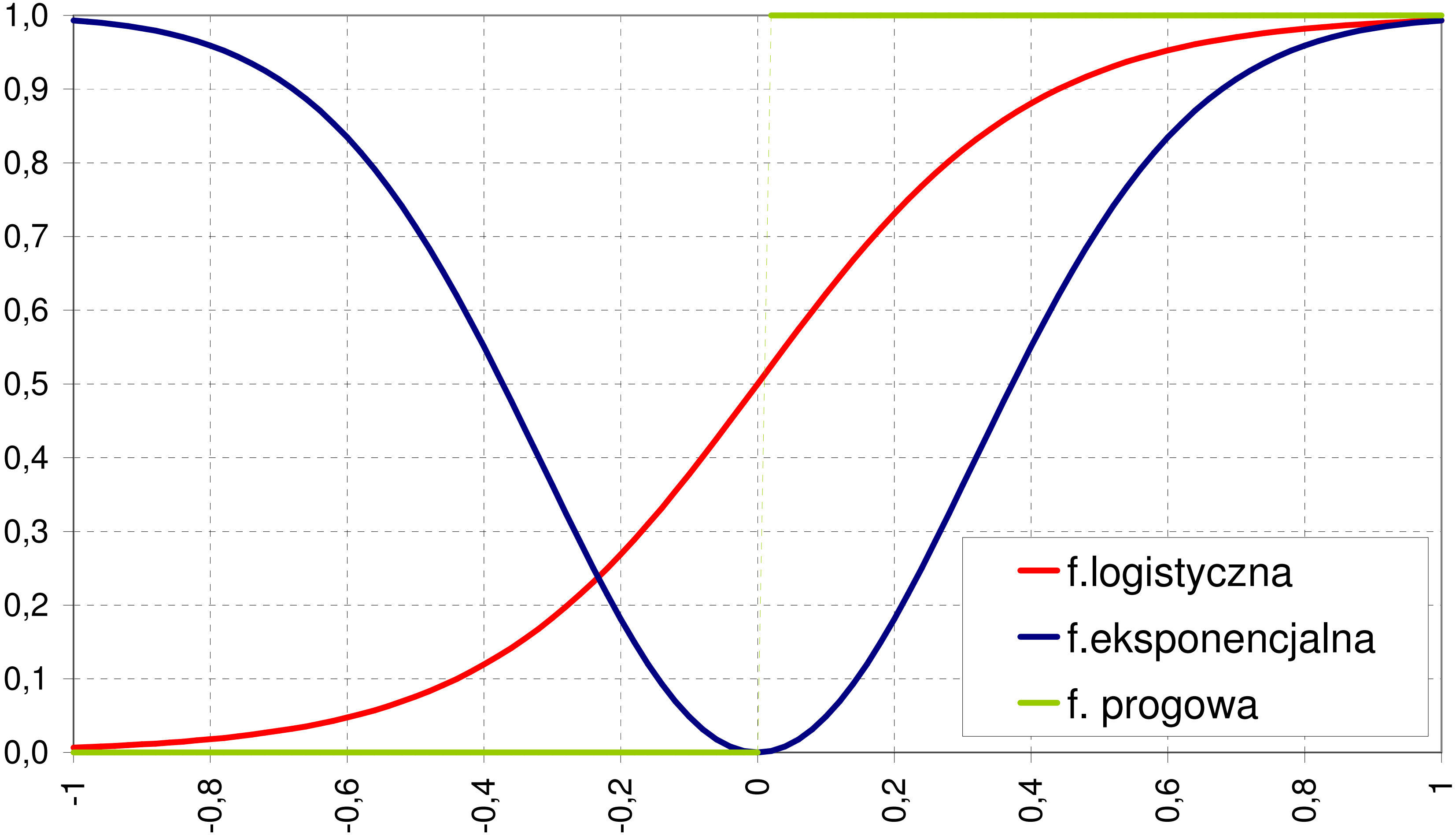
Funkcja przejścia

- G – funkcja eksponencjalna

$$G(z_t;\gamma,c)=1-\exp(-\gamma(z_t-c)^2)$$

- gdy $\gamma=0$, to model liniowy
- gdy $\gamma\to\infty$, to model liniowy

Funkcja przejścia



$$\gamma = 5 \quad c = 0$$

Estymacja

Teräsvirta (1994) „*conditional least squares*”:

$$y_t = g(\boldsymbol{\alpha}, F_{t-1}) + \varepsilon_t \qquad F_t = \{y_t, y_{t-1}, \ldots, y_1\}$$

$$E(\varepsilon_t \mid F_{t-1}) = 0$$

$$\mathrm{var}(\varepsilon_t \mid F_{t-1}) = \sigma^2$$

$$Q_T(\boldsymbol{\alpha}) = \sum_{t=1}^{T} [y_t - g(\boldsymbol{\alpha}, F_{t-1})]^2$$

Estymacja

- Estymator zgodny i asymptotycznie normalny

- Problemy techniczne metody gradientowej:
  - ESTR: γ silnie skorelowane z parametrami ϕ_2 (bez stałej)

1. standardyzuj wykładnik w G przez podzielenie go przez wariancję y
2. ustal startową wartość γ, np. $\gamma \approx 1$
3. jeśli algorytm gradientowy nie zbiega, to „grid search” - γ

Estymacja

- Estymator zgodny i asymptotycznie normalny

- Problemy techniczne metody gradientowej:
  – LSTR: γ, c do oszacowania i gdy $\gamma \to \infty$, to model TR (duży błąd oszacowania γ)

1. Skaluj parametry startowe (zmniejsz γ i zwiększ c [??]), podziel wykładnik w G przez odchylenie stand. y
2. Jeśli algorytm gradientowy nie zbiega, to „grid search”- c

Estymacja

- Możliwa estymacja MNK pod warunkiem, że znane γ, c

$$\hat{\phi}(\gamma, c) = \left(\sum_{t=1}^{T} x_t(\gamma, c) x_t(\gamma, c)' \right)^{-1} \left(\sum_{t=1}^{T} x_t(\gamma, c) y_t \right)$$

$$x_t(\gamma, c) \; = \; (x_t'(1 - G(s_t; \gamma, c)), x_t' G(s_t; \gamma, c))'$$

Specyfikacja

- wybór modelu liniowego (np. AR(p))
- testowanie liniowości modelu (przeciw STR) dla różnych zmiennych przejścia (*transition variables*)

  \+ wybór optymalnej zmiennej przejścia
- wybór między LSTR i ESTR

Testowanie modelu STR

- Hipotezy $H_0: \phi_1 = \phi_2$
  $H_1: \phi_1 \neq \phi_2$

- Problem z parametrami γ, c nieidentyfikowalnymi przy H_0 → niestandardowe rozkłady statystyk testowych

- Hipotezy $H_0: \gamma = 0$
  $H_1: \gamma \neq 0$

- Problem z parametrami c, ϕ_1, ϕ_2 nieidentyfikowalnymi przy H_0 → niestandardowe rozkłady statystyk testowych

Testowanie modelu STR

- Luukkonen, Saikkonen, Teräsvirta (1988):
  - Rozwinięcie modelu STR w szereg Taylora wokół $\gamma = 0$
  - Zastosowanie testu LM

Testowanie LSTR

- Szereg Taylora 1. rzędu:

$$y_t = \phi_1' x_t + (\phi_2 - \phi_1)' x_t G(z_t; \gamma, c) + \varepsilon_t$$

$$y_t = \beta_0' x_t + \beta_1' (x_t z_t) + e_t$$

$$e_t = \varepsilon_t + (\phi_2 - \phi_1)' x_t R_1(z_t; \gamma, c)$$

- Przy H_0

$$e_t = \varepsilon_t \qquad R_1(z_t; \gamma, c) = 0$$

…dlatego test LM

Testowanie LSTR c.d.

- Testowanie $\phi_1 = \phi_2$ ($\gamma = 0$) równoważne z

  $H_0 : \beta_1 = 0$

  $H_1 : \beta_1 \neq 0$

- Standardowy test LM (szczegóły później)
  - Nazywany tutaj: „*LM-type* test”

- Problem: kiedy $z_t = x_{i,t}$, trzeba usunąć $\beta_{1,0} z_t$ z „testowego” modelu regresji (współliniowość)
  - test nie nadaje się do testowania zmian stałej

Testowanie LSTR c.d.

- Rozwiązanie: szereg Taylora 3. rzędu

  $$y_t = \beta_0' x_t + \beta_1'(x_t z_t) + \beta_2'(x_t z_t^2) + \beta_3'(x_t z_t^3) + e_t$$

  bez $\beta_{1,0} z_t$, $\beta_{2,0} z_t^2$, $\beta_{3,0} z_t^3$

  lub uproszczona wersja:

  $$y_t = \beta_0' x_t + \beta_1'(x_t z_t) + \beta_2 z_t^2 + \beta_3 z_t^3 + e_t$$

- Test LM $\quad H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

Testowanie ESTR

- Rozwinięcie ESTR w szereg Taylora 1. rzędu:

$$y_t = \beta_0' x_t + \beta_1'(x_t z_t) + \beta_2'(x_t z_t^2) + e_t$$

$$e_t = \varepsilon_t + (\phi_2 - \phi_1)' x_t R_1(z_t; \gamma, c)$$

- lub uwzględniając 2 punkty przegięcia w ESTR – rozwinięcie 2. rzędu:

$$y_t = \beta_0' x_t + \beta_1'(x_t z_t) + \beta_2'(x_t z_t^2) + \beta_3'(x_t z_t^3) + \beta_4'(x_t z_t^4) + e_t$$

Testowanie ESTR c.d.

- Test typu LM

$$H_0 : \beta_1 = \beta_2 = \left(\beta_3 = \beta_4\right) = 0$$

Obliczanie statystyk LM

- Oszacuj model przy założeniu H_0

$$y_t = \hat{\beta}_0' x_t + \hat{e}_t \qquad RSS_0 = \sum_{t=1}^{T} \hat{e}_t^2$$

- Oszacuj „testową” regresję: $y \leftarrow x, xz$

$$RSS_1 = \sum_{t=1}^{T} \tilde{e}_t^2$$

- Statystyka: $LM = \dfrac{T(RSS_0 - RSS_1)}{RSS_0} \sim \chi^2_{(l.dodatkowych\ parametrow)}$

lub w małych próbach:

$$LM_F = \frac{(RSS_0 - RSS_1)/(l.dodatkowych\ parametrow)}{RSS_1/(T - l.wszystkich\ parametrow)} \sim F$$

Autokorelacja składnika losowego

- Niech $F(x_t;\theta) = \phi_1' x_t [1 - G(z_t;\gamma,c)] + \phi_2' x_t G(z_t;\gamma,c)$

- Oszacuj model STR i oblicz reszty $\hat{\varepsilon}_t$
- Oblicz $\nabla F(x_t;\hat{\theta}) = \partial F(x_t;\hat{\theta}) \big/ \partial \theta$ gdzie

$$\theta = [\phi_1 \quad \phi_2 \quad \gamma \quad c]'$$

- Oszacuj model liniowy i oblicz R-kwadrat:

$$\hat{\varepsilon}_t \leftarrow \nabla F(x_t;\hat{\theta}), \hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-2}, ..., \hat{\varepsilon}_{t-q}$$

- Rozszerzenie testu Godfreya (1979):

H_0: *brak autokorelacji* $LM = TR^2 \overset{a}{\sim} \chi_q^2$

Test pozostałej nieliniowości

- Model rozszerzony:

$$y_t = \phi_1' x_t + (\phi_2 - \phi_1)' x_t G_1(s_t; \gamma_1, c_1) + (\phi_3 - \phi_2)' x_t G_2(s_t; \gamma_2, c_2) + \varepsilon_t$$

- H_0: $\gamma_2 = 0$ lub $\phi_3 = \phi_2$

- Zamień G_2 na rozwinięcie w szereg Taylora 3. rzędu:

$$y_t = \beta_0' x_t + (\phi_2 - \phi_1)' x_t G_1(s_t; \gamma_1, c_1) + \beta_1' x_t s_t + \beta_2' x_t s_t^2 + \beta_3' x_t s_t^3 + e_t.$$

- Po przekształceniu, H_0: $\beta_1 = \beta_2 = \beta_3 = 0$

Test pozostałej nieliniowości

- Niech $F(x_t;\theta) = \phi_1' x_t [1 - G(z_t;\gamma,c)] + \phi_2' x_t G(z_t;\gamma,c)$

- Oszacuj model STR i oblicz reszty $\hat{\varepsilon}_t$
- Oblicz $\nabla F(x_t;\hat{\theta}) = \partial F(x_t;\hat{\theta}) \big/ \partial\theta$ gdzie

$$\theta = [\phi_1 \quad \phi_2 \quad \gamma \quad c]'$$

- Oszacuj model liniowy i oblicz R-kwadrat:

$$\hat{\varepsilon}_t \leftarrow \nabla F(x_t;\hat{\theta}),\ x_t z_t,\ x_t z_t^2,\ x_t z_t^3$$

- Statystyka LM

$$LM = TR^2 \sim \chi^2_{(l.dodatkowych\ parametrow)}$$

Wybór funkcji przejścia

- Sekwencja testów (statystyki LM):

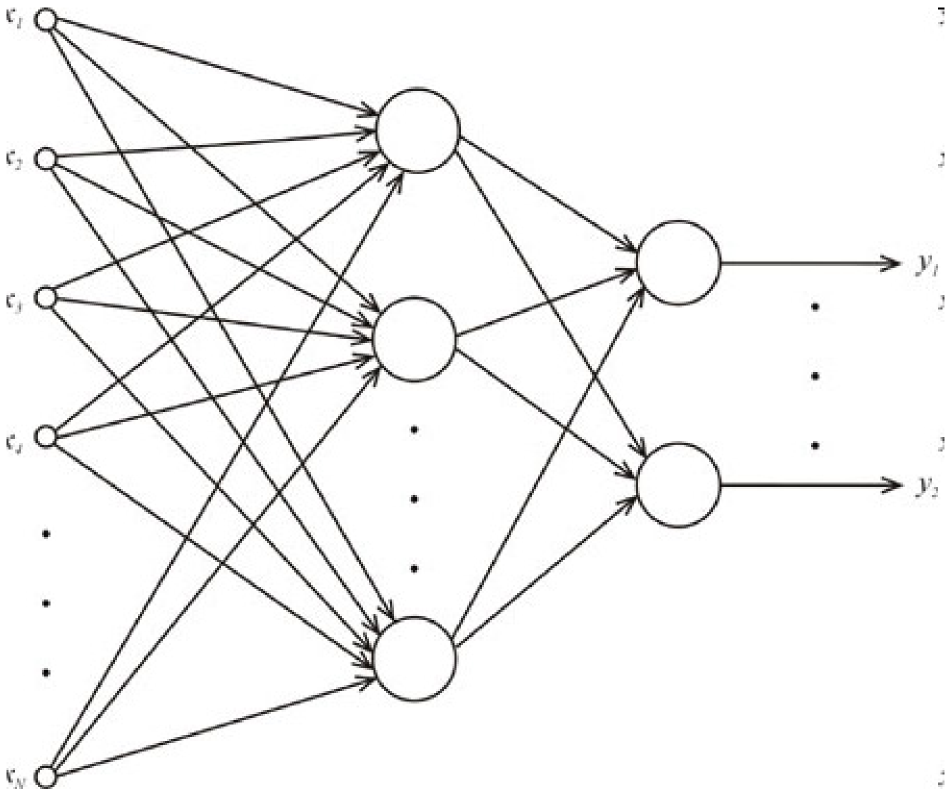
$$H_{01} : \beta_3 = 0$$

$$H_{02} : \beta_2 = 0 \big| \beta_3 = 0$$

$$H_{03} : \beta_1 = 0 \big| \beta_2 = 0, \beta_3 = 0$$

- Reguła decyzyjna:
  - Jeśli empiryczny poziom istotności (p-value) najmniejszy dla H_{02} , to wybierz model ESTR
  - Jeśli empiryczny poziom istotności najmniejszy dla H_{01} lub H_{03} , to wybierz model LSTR

Sieci neuronowe



Źródło: http://kik.pcz.czest.pl/nn/arch.php?art=3

Sieci neuronowe

- Model z jedną warstwą ukrytą

$$y_t = G(\mathbf{x}_t; \psi) + \varepsilon_t = \boldsymbol{\alpha}'\tilde{\mathbf{x}}_t + \sum_{i=1}^{h} \lambda_i F(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - \beta_i) + \varepsilon_t$$

$$\boldsymbol{\psi} = [\boldsymbol{\alpha}', \lambda_1, \ldots, \lambda_h, \tilde{\boldsymbol{\omega}}_1', \ldots, \tilde{\boldsymbol{\omega}}_h', \beta_1, \ldots, \beta_h]'$$

- Funkcja logistyczna F lub inna sigmoidalna

$$F(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - \beta_i) = (1 + e^{-(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - \beta_i)})^{-1}$$

$$\varepsilon_t \sim n.i.d.(0, \sigma^2)$$

Sieci neuronowe

- Dokładne dopasowanie modelu do danych
  - możliwe dowolnie dokładne przybliżenie funkcji ciągłej
  - nie proces generujący dane, ale model przybliżający prawdziwy proces
- Prognozowanie
- Ekonomiczna interpretacja zależności? Nie.

Identyfikowalność parametrów

- 3 problemy z identyfikowalnością
  - $h!$ permutacji neuronów (funkcji przejścia) ma taką samą wartość funkcji wiarygodności
  - dodatkowo: $F(x) = 1 - F(-x)$
  - duża liczba funkcji przejścia
- Rozwiązanie:
  - restrykcje: $\beta_1 \leq \ldots \leq \beta_h \quad \text{lub} \quad \lambda_1 \geq \ldots \geq \lambda_h$

$$\tilde{\omega}_{1i} > 0, i = 1, \ldots, h$$

  - odpowiednia specyfikacja modelu

Budowa modelu

- Wybór zmiennych
- Wybór liczby funkcji przejścia
  – wymaga estymacji modelu

Budowa modelu

- Wybór zmiennych
  - przybliżenie sieci neuronowej przez wielomian k-tego rzędu

$$G(\mathbf{x}_t;\boldsymbol{\psi}) = \boldsymbol{\pi}'\tilde{\mathbf{x}}_t + \sum_{j_1=1}^{q}\sum_{j_2=j_1}^{q}\theta_{j_1 j_2} x_{j_1,t} x_{j_2,t} + \cdots + \sum_{j_1=1}^{q}\cdots\sum_{j_k=j_{k-1}}^{q}\theta_{j_1\ldots j_k} x_{j_1,t}\cdots x_{j_k,t} + R(\mathbf{x}_t;\boldsymbol{\psi})$$

  - szacowanie modeli i optymalizacja kryterium informacyjnego: AIC, SBIC

Estymacja modelu

- Metoda Największej Wiarygodności
  - równoważna: Nieliniowa MNK

- Przydatna reparametryzacja

$$F\left(\gamma_i\left(\boldsymbol{\omega}_i'\mathbf{x}_t - c_i\right)\right) = \left(1 + e^{-\gamma_i\left(\boldsymbol{\omega}_i'\mathbf{x}_t - c_i\right)}\right)^{-1}$$

$$\gamma_i > 0,\ i = 1, \ldots, h$$

$$\|\boldsymbol{\omega}_i\| = 1 \qquad \omega_{i1} = \sqrt{1 - \sum_{j=2}^{q} \omega_{ij}^2} > 0, i = 1, \ldots, h$$

Estymacja modelu

- Wektor parametrów

$$\boldsymbol{\psi} = [\boldsymbol{\alpha}', \lambda_1, \ldots, \lambda_h, \gamma_1, \ldots, \gamma_h, \omega_{12}, \ldots, \omega_{1q}, \ldots, \omega_{h2}, \ldots, \omega_{hq}, c_1, \ldots, c_h]'$$

- Standaryzowanie zmiennych wejściowych:

  $Var(x)=1$

- Możliwość „koncentracji” funkcji wiarygodności – tzn. szacowanie parametrów w grupach

Estymacja c.d.

- Przyjmij za znane:

$$\boldsymbol{\phi} = [\gamma_1, \ldots, \gamma_h, \boldsymbol{\omega}_1', \ldots, \boldsymbol{\omega}_h', c_1, \ldots, c_h]'$$

- Zbuduj macierz Z dla regresji: $\mathbf{y} = \mathbf{Z}(\boldsymbol{\phi})\boldsymbol{\theta} + \varepsilon$

$$\mathbf{y}' = [y_1, y_2, \ldots, y_T],\ \varepsilon' = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T],\ \boldsymbol{\theta}' = [\boldsymbol{\alpha}', \lambda_1, \ldots, \lambda_h]$$

$$\mathbf{Z}(\boldsymbol{\phi}) = \begin{pmatrix} \tilde{\mathbf{x}}_1' & F(\gamma_1(\boldsymbol{\omega}_1'\mathbf{x}_1 - \mathrm{c}_1)) & \cdots & F(\gamma_h(\boldsymbol{\omega}_h'\mathbf{x}_1 - c_h)) \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\mathbf{x}}_T' & F(\gamma_1(\boldsymbol{\omega}_1'\mathbf{x}_T - \mathrm{c}_1)) & \cdots & F(\gamma_h(\boldsymbol{\omega}_h'\mathbf{x}_T - c_h)) \end{pmatrix}$$

Estymacja c.d.

- Szacuj parametry MNK:

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{Z}(\boldsymbol{\phi})'\mathbf{Z}(\boldsymbol{\phi})\right)^{-1}\mathbf{Z}(\boldsymbol{\phi})'\mathbf{y}$$

- Parametry $\boldsymbol{\phi} = [\gamma_1, \ldots, \gamma_h, \boldsymbol{\omega}'_1, \ldots, \boldsymbol{\omega}'_h, c_1, \ldots, c_h]'$ szacuj minimalizując sumę kwadratów reszt
  - algorytmy optymalizacji: BFGS, Levenberg-Marquardt

Wybór liczby funkcji przejścia

- Metoda „od małego do dużego”
  - dodawanie neuronów (*hidden units*)

- Testowanie czy $h+1$ neuron zbędny

$$y_t = \boldsymbol{\alpha}'\tilde{\mathrm{x}}_t + \sum_{i=1}^{h} \lambda_i F(\gamma_i(\boldsymbol{\omega}_i'\mathrm{x}_t - c_i)) + \lambda_{h+1} F(\gamma_{h+1}(\boldsymbol{\omega}_{h+1}'\mathrm{x}_t - c_{h+1})) + \varepsilon_t$$

$$\mathrm{H}_0 : \gamma_{h+1} = 0$$

Testowanie funkcji przejścia

- Rozwinięcie modelu w szereg Taylora 3. rzędu:

$$y_t = \boldsymbol{\pi}'\tilde{\mathbf{x}}_t + \sum_{i=1}^{h} \lambda_i F(\gamma_i(\boldsymbol{\omega}_i'\mathbf{x}_t - c_i)) + \sum_{i=1}^{q}\sum_{j=i}^{q} \theta_{ij} x_{i,t} x_{j,t} + \sum_{i=1}^{q}\sum_{j=i}^{q}\sum_{k=j}^{q} \theta_{ijk} x_{i,t} x_{j,t} x_{k,t} + \varepsilon_t^*$$

- Oszacuj model z h neuronami i oblicz reszty $\hat{\varepsilon}_t$
- Wyznacz wektor „score” $\hat{\mathbf{h}}_t = \left.\frac{\partial G(\mathrm{x}_t;\psi)}{\partial \psi'}\right|_{\psi=\hat{\psi}}$
  - Jeśli $\hat{\varepsilon}_t$ i $\hat{\mathbf{h}}_t$ nie są ortogonalne, to oszacuj regresję tych zmiennych i wyznacz reszty $\tilde{\varepsilon}_t$

Testowanie c.d.

- Oblicz $SSR_0 = \sum_{t=1}^{T} \tilde{\varepsilon}_t^2$
- Oszacuj regresję $\tilde{\varepsilon}_t$ na $\hat{\mathbf{h}}_t$ i v_t

$$\boldsymbol{v}_t = [x_{1,t}^2, x_{1,t}x_{2,t}, \ldots, x_{i,t}x_{j,t}, \ldots, x_{1,t}^3, \ldots, x_{i,t}x_{j,t}x_{k,t}, \ldots, x_{h,t}^3]$$

$$SSR_1 = \sum_{t=1}^{T} \hat{v}_t^2$$

- Oblicz reszty oraz

Testowanie c.d.

- Statystyka:

$$LM_{\chi^2}^{hn} = T\frac{SSR_0 - SSR_1}{SSR_0}$$

z m stopniami swobody

$$m = q(q+1)/2 + q(q+1)(q+2)/6$$

- W małych próbach:

$$LM_F^{hn} = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T-n-m)}$$

ma w przybliżeniu rozkład $F(m,T\text{-}n\text{-}m)$

$$n = (q+2)h + p + 1$$

Ewaluacja modelu

- Testy aukorelacji, niestabilności parametrów
- Analiza prognoz