Econometrics Spring School 2016
Econometric Modelling

Jurgen A Doornik, David F. Hendry, and Felix Pretis

George-Washington University

March 2016

Lecture 9: Non-linear Model Selection

Motivation:

- Non-linearity inherent in economics (and many other fields)
- Linear representation is a simplifying assumption
- If incorrect, model will be mis-specified

Objective:

- Test for evidence of non-linearity
- Commence with general non-linear approximation.
  Investigate large class of functions: polynomials one possibility:
    - needs to be identified and congruent;
    - approximate wide range of non-linear models;
    - maintains linearity in the parameters;
    - easy to orthogonalize.
- But non-generalized polynomials not invariant to transformations

Establish operating characteristics

Non-linear model selection:

- Testing for non-linearity;
  - Caveat: ...like going to the zoo to look at non-elephants[1]
- Mimicking a near-orthogonal representation;
- Avoiding extreme observations leading to non-normality – impulse-indicator saturation and step-indicator saturation;
- Preventing excess retention of irrelevant variables – choice of significance level.

A successful algorithm requires the synthesis of all developments to be implemented

[1] With thanks to Anders Rahbek.

Assume LDGP is:

$$y_t = f\left(z_{1,t}, \ldots, z_{k,t}; \theta\right) + \epsilon_t \quad \text{where} \quad \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{1}$$

for $t = 1, \ldots, T$, with $\theta \in \boldsymbol{\Theta} \subseteq \mathbb{R}^k$.

Problems include:

- specification of functional form, $f\left(\cdot\right)$;
- identification of θ;
- selection of relevant variables, $\mathbf{z}_t' = \left(z_{1,t}, \ldots z_{k,t}\right)$ from available candidates $\left(z_{1,t}, \ldots z_{K,t}\right)$ where $K \geq k$.

Specify a GUM that nests the LDGP to ensure the initial formulation is congruent:

$$y_t = \sum_{j=1}^{K} \beta_j \sum_{p=1}^{P} h_p \left(z_{1,t}, \ldots z_{K,t} \right) + \nu_t \quad \text{where} \quad \nu_t \sim \mathsf{IN}\left[0, \sigma_\nu^2\right] \qquad (2)$$

k relevant and $K - k$ irrelevant variables.
P approximation bases.
Test for non-linearity to see if it is viable to reduce to:

$$y_t = \sum_{j=1}^{K} \beta_j z_{j,t} + \nu_t. \qquad (3)$$

If do not reject, proceed with linear GUM.
If reject, non-linearity is established.

2 cases:

1. **GUM nests LDGP**
   Approximation is exact, $f(\cdot) \subseteq h_p(\cdot)$
   Congruency is meaningful and testable, even with more variables than observations
   Consistent model selection: $\alpha \to 0$ as $T \to \infty$ so irrelevant variables eventually retained with prob 0.
   Functional form concerns whether more parsimonious representation found.
2. **GUM doesn't nest LDGP but is approximation**
   Not clear what consistency means – test of approx.
   Cannot prove consistency as non-nested LDGP.
   Consistency feasible with encompassing test stage.
   Congruency still operational.

Commence with general approximation $h_p(\mathbf{z}_i)$:

$$y_t = \beta'\mathbf{z}_t + \gamma'\mathbf{g}(\mathbf{w}_t) + \sum_{t=1}^{T} \delta_t \mathbf{1}_{\{t=t_i\}} + v_t \quad t = 1, \ldots, T \tag{4}$$

where $v_t \sim \mathsf{NID}\left(0, \sigma_v^2\right)$.
$\mathbf{z}_t = (z_{1,t}, \ldots, z_{K,t})'$ is $(K \times 1)$ vector of potentially relevant variables.
$\mathbf{g}(\mathbf{w}_t)$ is $(M \times 1)$ vector of non-linear transformations on standardized potentially relevant variables or factors, $\mathbf{w}_t$:

$$w_{j,t} = \frac{z_{j,t} - \overline{z}_j}{\sigma_{z_j}} \qquad j = 1, \ldots, K, \quad \text{or} \quad \mathbf{w}_t = \mathbf{\Lambda}^{1/2}\mathbf{H}'(\mathbf{z}_t - \overline{\mathbf{z}}) \tag{5}$$

$\sum_{t=1}^{T} \mathbf{1}_{\{t=t_i\}}$ is a set of saturating indicators.

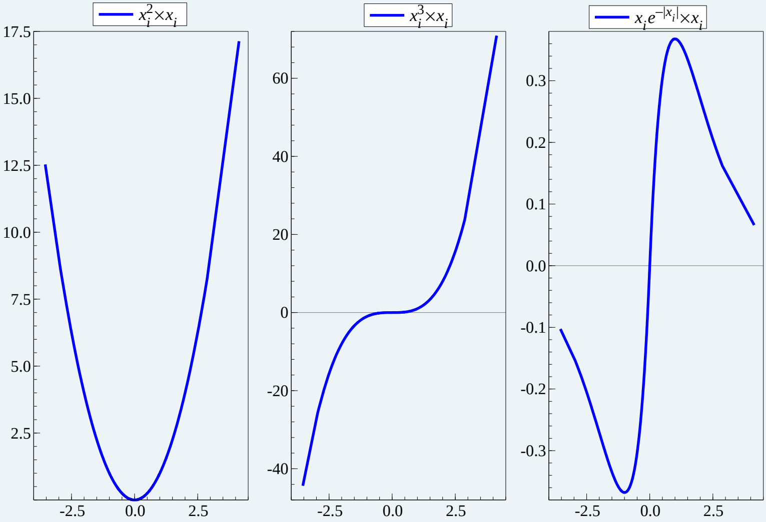
Key question: What $\mathbf{g}(\mathbf{w}_t)$ specification?

Many possible approximations:

- Polynomials, Hermite, Chebyshev, ...
- Fourier series approximations
- Asymptotic series
- Logistic functions

Range of different bases – different shapes of functions.
Ability to parsimoniously approximate depends on LDGP.
Suggests many bases but only first few terms of each.
Preferable to one base with longer approximation.
Issue: approximations not orthogonal – could 'cancel', hence selection.

$$\begin{aligned}\gamma'\mathbf{g}(\mathbf{w}_t) &= \sum_{j=1}^{K}\sum_{l=j}^{K}\beta_{jl}w_{j,t}w_{l,t} && \text{[2}^{\text{nd}}\text{ order polynomials]} \\ &+ \sum_{j=1}^{K}\sum_{l=j}^{K}\sum_{q=l}^{K}\beta_{jlq}w_{j,t}w_{l,t}w_{q,t} && \text{[3}^{\text{rd}}\text{ order polynomials]} \\ &+ \sum_{j=1}^{K}\alpha_{jj}\left\{w_{j,t}e^{-|w_{j,t}|}\right\} && \text{[exponentials]} \qquad (6)\end{aligned}$$


$x_i^2 \times x_i$
17.5
15.0
12.5
10.0
7.5
5.0
2.5
-2.5
0.0
2.5
$x_i^3 \times x_i$
60
40
20
0
-20
-40
-2.5
0.0
2.5
$x_i e^{-|x_i|} \times x_i$
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-2.5
0.0
2.5

Selection undertaken on (11):

$$\widehat{y}_t = \sum_{j=1}^{k^*} \widehat{\beta}_j z_{j,t} + \sum_{j=1}^{m^*} \widehat{\gamma}_j g_j\left(\mathbf{w}_t\right) + \sum_{i=1}^{q} \widehat{\delta}_i \mathbf{1}_{\{t=t_i\}} \tag{7}$$

k^* = no. linear regressors retained;
m^* = no. non-linear transformations retained;
q = indicators retained.
Final stage – test approximation (7) against preferred functional form, $\psi\left(\mathbf{z}_t\right)$ (e.g. LSTAR, theory-motivated etc).
Encompassing test: $H_0 : \gamma_j = 0, \forall j$:

$$y_t = \sum_{j=1}^{k^*} \widehat{\beta}_j z_{j,t} + \sum_{j=1}^{m^*} \gamma_j g\left(\mathbf{w}_t\right) + \sum_{i=1}^{q} \delta_i \mathbf{1}_{\{t=t_i\}} + \lambda' \psi\left(\mathbf{z}_t\right) + \eta_t. \tag{8}$$

We shall examine each aspect in turn before undertaking selection jointly.
In practice, all aspects should be implemented jointly.

1. Test for non-linearity
2. Outlier detection
3. Non-linear functions
4. Joint selection using a parsimonious non-linear function
5. Use of SIS to obtain an encompassing model

Test for non-linearity in general linear model by low-dimensional portmanteau test in Castle and Hendry (2010) uses cubics of principal components $\boldsymbol{w}_t$ of the $\boldsymbol{z}_t$.

Let $\boldsymbol{z}_t \sim \mathsf{D}_n[\boldsymbol{\mu}, \boldsymbol{\Omega}]$, where $\boldsymbol{\Omega} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'$ with $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_n$.

Then $\boldsymbol{w}_t^* = \boldsymbol{H}'\boldsymbol{z}_t \Rightarrow \boldsymbol{w}_t^* \sim \mathsf{D}_n[\boldsymbol{H}'\boldsymbol{\mu}, \boldsymbol{\Lambda}]$. Empirically:

$\widehat{\boldsymbol{\Omega}} = T^{-1}\sum_{t=1}^{T}(\boldsymbol{z}_t - \overline{\boldsymbol{z}})(\boldsymbol{z}_t - \overline{\boldsymbol{z}})' = \widehat{\boldsymbol{H}}\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{H}}'$ so that $\boldsymbol{w}_t = \widehat{\boldsymbol{\Lambda}}^{-1/2}\widehat{\boldsymbol{H}}'(\boldsymbol{z}_t - \overline{\boldsymbol{z}})$ leading to $\boldsymbol{w}_t \underset{\text{app}}{\sim} \mathsf{D}_n[\boldsymbol{0}, \boldsymbol{I}]$.

Test for non-linearity in general linear model by low-dimensional portmanteau test in Castle and Hendry (2010) uses cubics of principal components $\boldsymbol{w}_t$ of the $\boldsymbol{z}_t$.

Let $\boldsymbol{z}_t \sim \mathsf{D}_n[\boldsymbol{\mu}, \boldsymbol{\Omega}]$, where $\boldsymbol{\Omega} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'$ with $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_n$.

Then $\boldsymbol{w}_t^* = \boldsymbol{H}'\boldsymbol{z}_t \Rightarrow \boldsymbol{w}_t^* \sim \mathsf{D}_n[\boldsymbol{H}'\boldsymbol{\mu}, \boldsymbol{\Lambda}]$. Empirically:

$\widehat{\boldsymbol{\Omega}} = T^{-1}\sum_{t=1}^{T}(\boldsymbol{z}_t - \overline{\boldsymbol{z}})(\boldsymbol{z}_t - \overline{\boldsymbol{z}})' = \widehat{\boldsymbol{H}}\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{H}}'$ so that $\boldsymbol{w}_t = \widehat{\boldsymbol{\Lambda}}^{-1/2}\widehat{\boldsymbol{H}}'(\boldsymbol{z}_t - \overline{\boldsymbol{z}})$ leading to $\boldsymbol{w}_t \underset{\mathsf{app}}{\sim} \mathsf{D}_n[\boldsymbol{0}, \boldsymbol{I}]$.

If test rejects, create $\boldsymbol{g}(\boldsymbol{w}_t)$, otherwise $\boldsymbol{g}(\boldsymbol{z}_t) = \boldsymbol{z}_t$: presently, implemented general cubics with exponential functions.
$u_{1,i,t} = w_{i,t}^2;\ u_{2,i,t} = w_{i,t}^3;\ u_{3,i,t} = w_{i,t}e^{-|w_{i,t}|}$.

When $\boldsymbol{\Omega}$ is non-diagonal, each $w_{i,t}$ is a linear combination of every $z_{i,t}$, so $w_{i,t}^2$ involves squares and cross-products of every $z_{i,t}$ etc.

Advantages of test are:

- Low dimensionality–may entail more non-linear functions than observations;
- No collinearity between elements of $\boldsymbol{w}_t$;
- Includes most important sources of departure from linearity, e.g. asymmetry.

Number of potential regressors for cubic polynomials is:

$$M_K = K\left(K+1\right)\left(K+5\right)/6.$$

Explosion in number of terms as $K = r \times (s+1)$ increases:

K	1	2	3	4	5	10	15	20	30	40
M_K	3	9	19	30	55	285	679	1539	5455	12300

Quickly reach huge M_K: **but only** $3K$ **if use** $w^k_{i,t-j}$.
Later address perfect collinearity between $\boldsymbol{z}_t$ and $\boldsymbol{w}_t$.

Non-Linear Model Selection in Practice

- Load data "**nonlin_nobreak_example.in7**"
- Batch file: "**nonlinear_data_example.fl**"
- Artificial data with non-linear DGP:

$$\text{DGP: } y_t = \beta_1 x_{1,t} + \beta_2 x_{1,t}^2 + \beta_3 x_{1,t}^3 + \beta_4 x_{2,t}^2 + \epsilon_t \tag{9}$$

where $\mathsf{E}[\psi_i] = t_{\beta_i} = 3$ for $i = 1, \ldots, 4$, and $\epsilon_t \sim N(0, 1)$.

- **Estimate the mis-specified model**:

$$y_t = \beta_0 + \sum_{i=1}^{10} \beta_i x_{i,t} \tag{10}$$

The non-linearity index test applied to a linear model of y_t, where the regressors include an intercept, $x_{1,t}$ to $x_{10,t}$:

The test is significant at $\mathsf{p} = 0.008$ with $\mathsf{F}(30, 59) = 2.10$.

Specify GUM $\rightarrow$ Estimate model.
Test $\rightarrow$ Index test for non-linearity.
Quadratic, cubic and exponential principal components included.

Caveat: model also fails other diagnostics – must jointly test for outliers using IIS/SIS (Castle, Doornik, Hendry, and Pretis (2015)).

- K=10 variables → 285 regressors if all non-linear combinations (highly correlated)
- Principal components: low dimensionality

Create principal components of set of explanatory variables:

$$(x_{1,t}, \ldots, x_{10,t})$$

Other models → Descriptive statistics using PcGive.
Select regressors and choose Principal component analysis. Make sure you select the Save components in database option.

Compute squares and cubics of principal components using calculator/algebra:

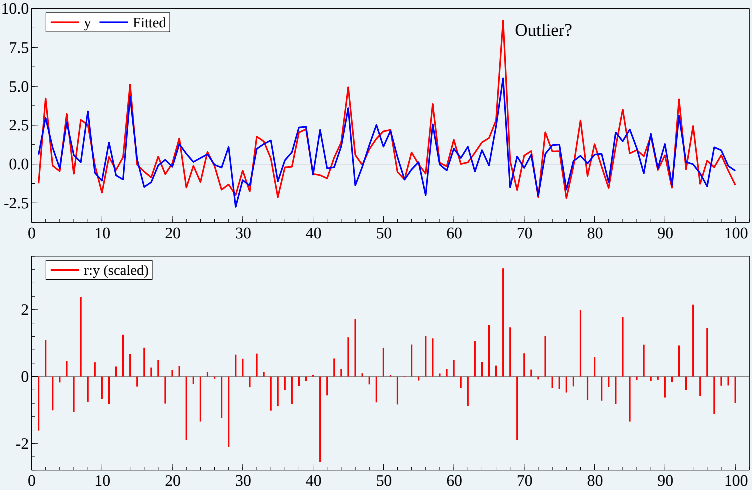
- `PC1sq = PC1^2;`
- `PC1cub = PC1^3;`
- `PC1exp = PC1*exp(-abs(PC1));`
- ...

Specify GUM (up to PC3) and select at tight significance level, $p_\alpha = 0.01$

Model Selection using non-linear transformations of principal components:

	Coefficient	Std.Error	t-value
x1	1.50656	0.1235	12.2
PC1_sq	0.160795	0.04795	3.35
PC3_sq	0.147337	0.05937	2.48

```
AR 1-2 test:      F(2,95)   =  0.77799 [0.4622]
ARCH 1-1 test:    F(1,98)   = 0.027482 [0.8687]
Normality test:   Chi^2(2)  =   3.6125 [0.1643]
Hetero test:      F(6,93)   =   2.8462 [0.0137]*
Hetero-X test:    F(9,90)   =   5.8234 [0.0000]**
RESET23 test:     F(2,95)   =   13.508 [0.0000]**
```



Must tackle non-linearities and shifts jointly!

Extend GUM to include Step-Indicators, select at $p_{\alpha=0.01}$:

$$y_t = \beta' \mathbf{z}_t + \gamma' \mathbf{g}(\mathbf{w}_t) + \sum_{t=1}^{T} \delta_t \mathbf{1}_{\{t \leq t_i\}} + v_t \quad t = 1, \ldots, T \qquad (11)$$

Results:

	Coefficient	Std.Error	t-value	t-prob
x1	1.34988	0.1223	11.0	0.0000
PC1_sq	0.142090	0.04767	2.98	0.0036
S1:66	-4.94129	1.138	-4.34	0.0000
S1:67	5.15142	1.145	4.50	0.0000

```
AR 1-2 test:       F(2,94)    =   1.3612 [0.2614]
ARCH 1-1 test:     F(1,98)    =  0.29562 [0.5879]
Normality test:    Chi^2(2)   =   1.0758 [0.5840]
Hetero test:       F(5,93)    =  0.54736 [0.7399]
Hetero-X test:     F(6,92)    =  0.58405 [0.7422]
RESET23 test:      F(2,94)    =   4.0001 [0.0215]*
```

Tackling non-linearities and shifts jointly.

- **So far**: non-linear DGP approximated by non-linear principal component + step-shifts.
- **Now**: test against preferred functional form
- Final stage – test approximation against preferred functional form, $\psi(\mathbf{z}_t)$ (e.g. LSTAR, theory-motivated etc).

Encompassing test: $H_0 : \gamma_j = 0, \forall j$:

$$y_t = \sum_{j=1}^{k^*} \widehat{\beta}_j x_{j,t} + \sum_{j=1}^{m^*} \gamma_j g(\mathbf{w}_t) + \sum_{i=1}^{q} \delta_i \mathbf{1}_{\{t \leq t_i\}} + \lambda' \psi(\mathbf{x}_t) + \eta_t.$$

Use 'theory motivated' specification:

$$\lambda' \psi(\mathbf{x}_t) = \lambda_1 x_{1,t}^2 + \lambda_2 x_{1,t}^3 + \lambda_3 x_{2,t}^2$$

Encompassing Model:

$$y_t = \beta_1 x_{1,t} + \gamma_1 w_{1,t}^2 + \delta_1 S_{t=66} + \delta_2 S_{t=67} + \lambda_1 x_{1,t}^2 + \lambda_2 x_{1,t}^3 + \lambda_3 x_{2,t}^2$$

```
Test for excluding:
[0] = PC1_sq
[1] = S1:66
[2] = S1:67
Subset F(3,93)   =  0.49576 [0.6861]
```

Detected shifts drop out when non-linear functional form specified.

Effect of using SIS when selecting from GUM nesting a non-linear DGP

Retention at 1% (x's selected over)		
	without SIS	with SIS
x, $\psi = 3$	0.68	0.63
x^2, $\psi = 3$	0.68	0.56
x^3, $\psi = 3$	0.67	0.62
SIS gauge	-	0.03

Little effect when DGP is truly non-linear.

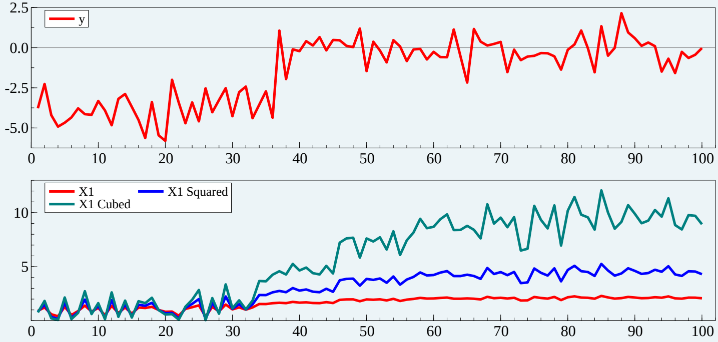
Non-Linear Example 2: Breaks

So far:

- Automatic non-linear extensions
- SIS steps not retained if DGP truly non-linear

Now:

- Spurious non-linearity due to shifts
- Load dataset: "nonlin_break_example.in7"



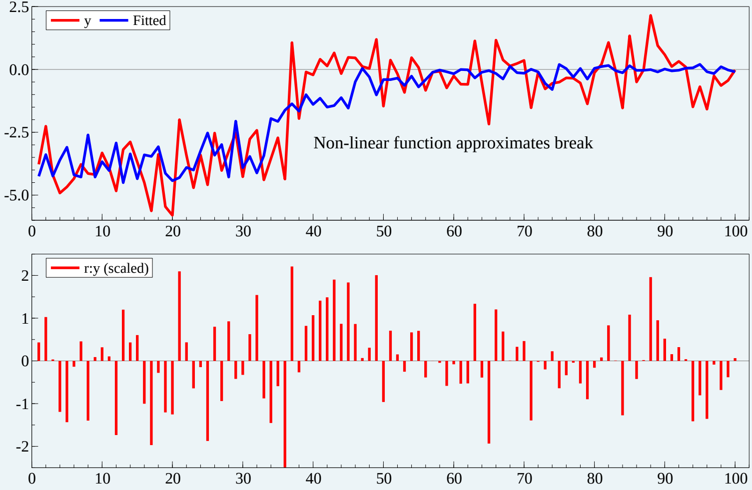
Select model at $p_\alpha = 0.01$ starting from GUM:

$$y_t = \beta_1 x_{1,t} + \beta_2 x_{1,t}^2 + \beta_3 x_{1,t}^3 + \sum_{i=2}^{5} \lambda_i x_{i,t}$$

Selection results:

	Coefficient	Std.Error	t-value	t-prob
x1	-13.4025	1.139	-11.8	0.0000
x3	0.0879400	0.1070	0.822	0.4131
x1_sq	11.9787	1.437	8.33	0.0000
x1_cub	-2.66853	0.4369	-6.11	0.0000

```
AR 1-2 test:        F(2,94)    =   0.80101 [0.4519]
ARCH 1-1 test:      F(1,98)    =    1.0423 [0.3098]
Normality test:     Chi^2(2)   =0.0010385 [0.9995]
Hetero test:        F(7,92)    =    2.6783 [0.0144]*
Hetero-X test:      F(11,88)   =    1.6932 [0.0880]
RESET23 test:       F(2,94)    =    10.153 [0.0001]**
```



Spurious non-linearity?

Start with GUM including step-indicators:

$$y_t = \beta_1 x_{1,t} + \beta_2 x_{1,t}^2 + \beta_3 x_{1,t}^3 + \sum_{i=2}^{5} \lambda_i x_{i,t} + \sum_{i=1}^{q} \delta_i \mathbf{1}_{\{t \leq t_i\}}$$

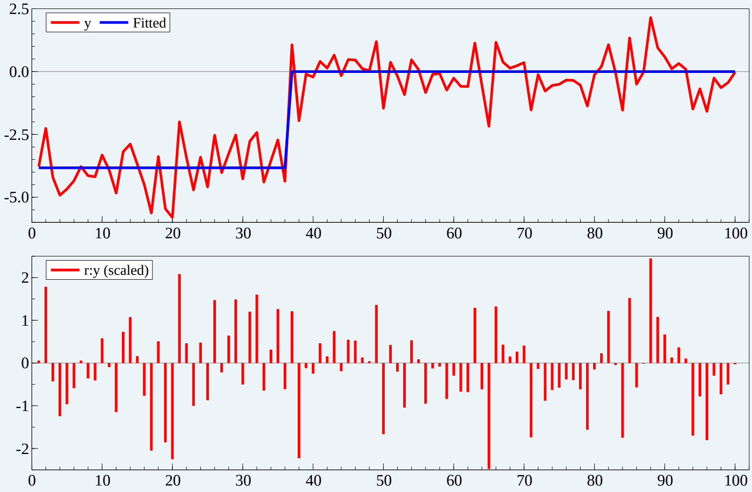
Yields:

	Coefficient	Std.Error	t-value	t-prob
S1:36	-3.82795	0.1463	-26.2	0.0000

```
AR 1-2 test:        F(2,97)    = 0.060408 [0.9414]
ARCH 1-1 test:      F(1,98)    =  0.30641 [0.5812]
Normality test:     Chi^2(2)   =  0.18369 [0.9122]
RESET23 test:       F(1,98)    =6.4913e-028 [1.0000]
```

Which nearly coincides with the DGP:

$$y_t = \mu + \lambda 1_{t \geq 35} + \epsilon_t \qquad (12)$$



Step-shift identified, spurious non-linearity removed.

SIS when non-linear transformations spuriously approximate a break of magnitude λ_1 in a linear DGP:

	Retention at 1% (x's selected over)			
	$\lambda_1 = 2\sigma_\epsilon$		$\lambda_1 = 4\sigma_\epsilon$	
	without SIS	with SIS	without SIS	with SIS
x, $\psi = 0$	0.41	0.02	0.83	0.02
x^2, $\psi = 0$	0.66	0.02	0.96	0.02
x^3, $\psi = 0$	0.3	0.02	0.83	0.01
T_1 step	-	0.62	-	0.94
SIS gauge	-	0.02	-	0.02

General Approach – Model Discovery Extensions:

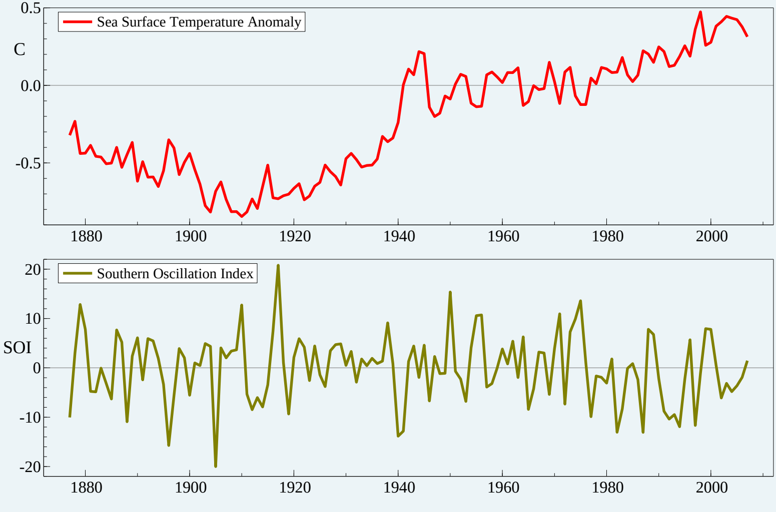
- Automatic non-linear extensions
- Jointly with indicator saturation

SIS as insurance mechanism:

- Non-linearities present: little effect of using SIS, steps not retained.
- Unknown shifts: identified through SIS rather than attributed to non-linearities

Global SST – Climate/Weather Indicator

Shape of trend? Breaks in series? El Niño Southern Oscillation?



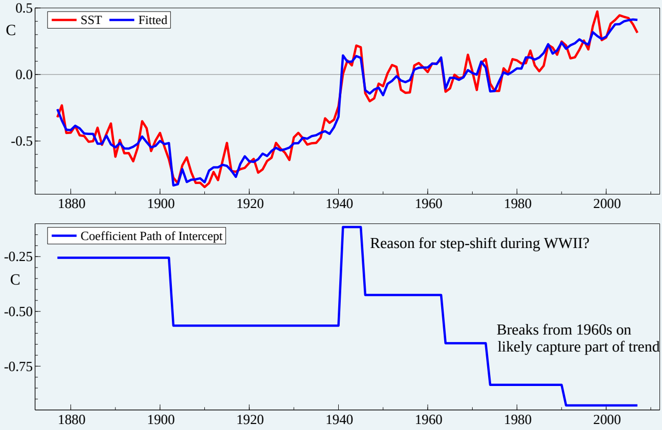
Model:

$$\mathbf{y_t} = \mathbf{f}(\mathbf{z_t}) + \beta \mathbf{x_t} + \delta \mathbf{D} + \epsilon_\mathbf{t} \quad (13)$$

- **T**=131 (1877-2007)
- **y**= Global Mean SST Anomalies (C) relative to 1950-79
- **z**=trend, **x**= Southern Oscillation Index (SOI) (atm. pressure at SL)
- **N** = 131 + 6 = 137 variables

Specification:

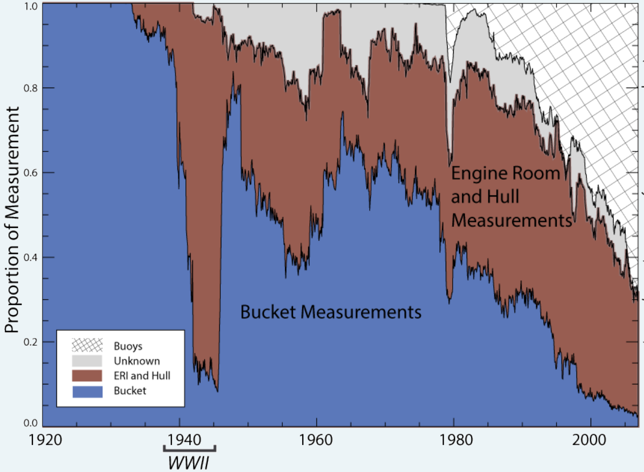
- SIS, $p_\alpha = 0.01$ (1%)
- Non-linear trend: B-spline basis (5 degree polynomial)



Multiple breaks, note:

- **1940**, $\hat{\delta}_1 = \mathbf{0.45^{***}C}$ (se=0.04)
- **1945**, $\hat{\delta}_2 = \mathbf{-0.31^{***}C}$ (se=0.04)

- WWII: **1941/1942 Measurements**: buckets to engine intake
  - Danger of measurements (light)
  - Americans joined 1941/1942
- Post-WWII: Partly changed back



(Adapted from Kennedy et al. 2012, *J. Geoph. Res.*)

- WWII: **1941/1942** **Measurements**: buckets to engine intake
  - Danger of measurements (light)
  - Americans joined 1941/1942
- Post-WWII: Partly changed back
- Buckets: **cold bias** ($\approx 0.3C$) (Matthews, 2012)
  - SIS: $\hat{\delta}_1 = 0.45$ (± 0.04)

SOI Effect (linear)

- SIS: $\hat{\beta} = -\mathbf{0.004}^{***}$ (se=0.001)
- Theory consistent: SOI > 0 (La Niña) $\rightarrow$ lower temp. (hiatus?!)

Breaks

- Could correct SST record for '**bucket bias**'

Non-linearities

- Non-linear transformations of trend retained

Automatic selection well developed for linear models

Algorithm can be extended to handle non-linear functions – not yet automated

- Test of functional form against linear
- Low order approximations for a range of bases
- Polynomials and exponentials $\times$ polynomials very general
- Indicator saturation to remove extreme observations/level shifts
- Number of potential non-linear variables large: choice of significance level
- Methods for more variables than observations
- Encompassing tests against specific functional forms

Complexity of empirical modeling – but theory consistent, empirically congruent model can be developed.

OXFORD MARTIN SCHOOL UNIVERSITY OF OXFORD Institute for New Economic Thinking AT THE OXFORD MARTIN SCHOOL

Joint modelling of dynamics, location shifts, relevant variables and non-linearities essential.
Automatic model selection despite $N > T$ seems a viable approach to tackling all complications jointly.

- IIS and SIS do not preclude finding non-linearites, and non-linearities removed indicators found in linear specifications.
- Not removing the large outliers could hide the presence of other variables, including the non-linearities.
- Indicator saturation: insurance mechanism in non-linear modelling.

Castle, J. L., J. A. Doornik, D. F. Hendry, and F. Pretis (2015).
Detecting location shifts during model selection by step-indicator saturation.
Econometrics 3(2), 240–264.

Castle, J. L. and D. F. Hendry (2010).
A low-dimension portmanteau test for non-linearity.
Journal of Econometrics 158(2), 231–245.