General-to-Specific Time Series Modelling

Felix Pretis

University of Oxford

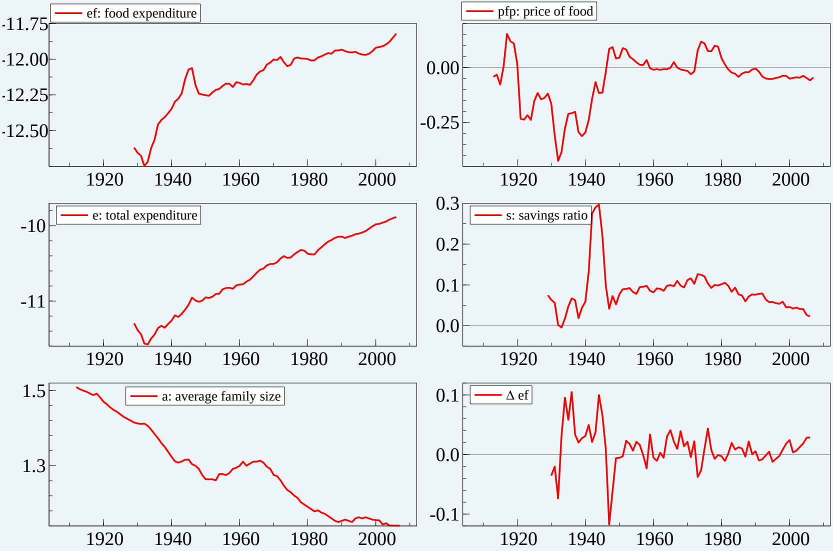
Michaelmas 2017

Lecture 2: Indicator Saturation

Core References for Lecture 2:

- Hendry, Johansen, and Santos (2008)* – Impulse Indicator Saturation (IIS)
- Castle, Doornik, Hendry, and Pretis (2015)* – Step Indicator Saturation (SIS)
- Johansen and Nielsen (2009) – IIS Asymptotic Theory
- Johansen and Nielsen (2016) – IIS Asymptotic Theory

US Food Demand (Expenditure) – Hendry and Mizon (2011):



US Food Demand (Expenditure) – Hendry and Mizon (2011):

- The data variables are (lower case denoting logs):
  - e_f is constant price, per capita, expenditure on food
  - e is constant price, per capita, total expenditure
  - p is deflator of total expenditure
  - y is constant price, per capita, income
  - $p_f - p$ is real price of food
  - $s = (y - e)$ is an approximation to the savings ratio
  - a is average family size–demographic effects
  - n is total population of the USA–
    should be irrelevant as per capita data.

There are considerable changes over the period:

- e_f and e fall sharply at the beginning of the Great Depression, rise substantially till WWII, fall after, then resume a gentle rise,
- so Δe_f is much more volatile pre WWII: Δe has a similar but less pronounced pattern).
- $p_f - p$ is quite volatile till after WWII, then is relatively stable,
- s rises from 'forced saving' in WWII.
- a has fallen considerably, partly reflecting changes in social mores.

- Tobin (1950) modelled US food demand:
  used time series 1912-48.
  We use extended time-series data, updated by Reade (2008).
- The basic theory is:

$$e_f = f\left(e, p_f - p, s, a\right) \tag{1}$$

- Conventional theory expects:

$$\frac{\partial e_f}{\partial e} > 0, \quad \frac{\partial e_f}{\partial \left(p_f - p\right)} < 0, \quad \frac{\partial e_f}{\partial a} < 0, \quad \frac{\partial e_f}{\partial n} = 0 \tag{2}$$

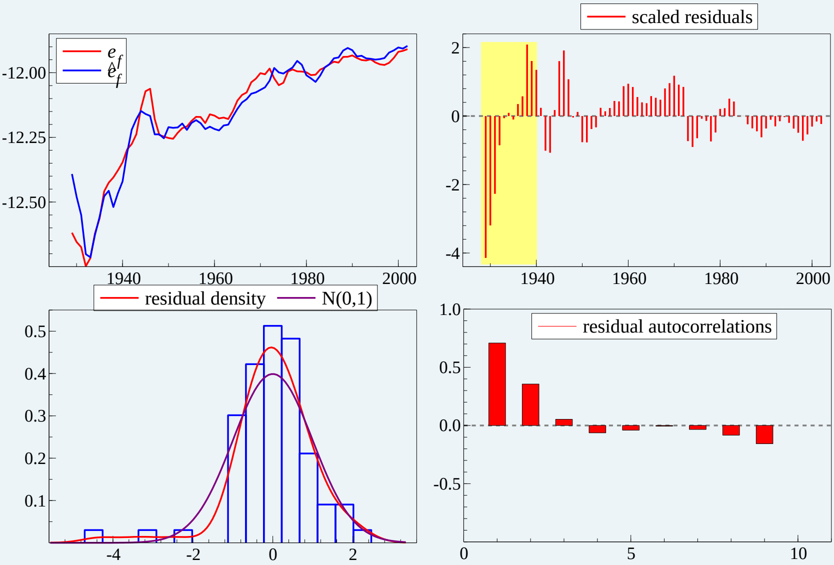
The static theory model estimates are:

$$e_{f,t} = \underset{(4.02)}{5.30} + \underset{(0.14)}{0.77}\, e_t + \underset{(0.08)}{0.11}\, (p_f - p)_t + \underset{(0.14)}{0.72}\, s_t - \underset{(0.23)}{0.36}\, a_t - \underset{(0.22)}{0.73}\, n_t$$

$$R^2 = 0.94 \quad \chi^2_{nd}(2) = 19.5^{**} \quad F_{arch}(1,72) = 216.8^{**} \quad F_{ar}(2,66) = 44.3^{**}$$
$$\widehat{\sigma} = 0.055 \quad F_{reset}(2,66) = 18.1^{**} \quad F_{het}(10,63) = 23.2^{**}$$

- The static economic-theory model has a very poor fit, and does not adequately capture behaviour of observed data.
- The price elasticity $(p_f - p)_t$ has the 'wrong sign', contradicting (2), but is insignificant.
- Although it is theoretically irrelevant, population n_t is significant.
- Finally, every mis-specification test strongly rejects.
  Next Figure shows the estimated model fails to describe the 1930s.

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



- Magnus and Morgan (1999) (‘competition’):
  Most contributors found dynamic models were non-constant over full sample 1931–1989, so modelled post 1950 only.

- Magnus and Morgan (1999) (‘competition’):
  Most contributors found dynamic models were non-constant over full sample 1931–1989, so modelled post 1950 only.
- Hendry (1999):
  i) Added Indicators for inter-war & post-war
  - Food program, Great Depression

  ii) Reversed procedure: indicators for all observations from 1950s onwards
  - 1970s
- Retained significant and re-select.

- Magnus and Morgan (1999) (‘competition’):
  Most contributors found dynamic models were non-constant over full sample 1931–1989, so modelled post 1950 only.
- Hendry (1999):
  i) Added Indicators for inter-war & post-war
  - Food program, Great Depression

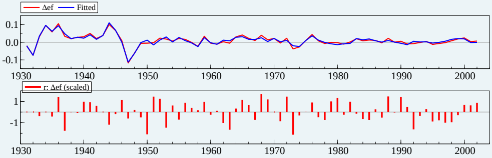
  ii) Reversed procedure: indicators for all observations from 1950s onwards
  - 1970s
- Retained significant and re-select.
- Hendry (1999) found a constant equation over 1931–1989 by adding impulse indicators pre-1950 for large outliers, identified as being due to a food program and post-war de-rationing.

$$
\begin{aligned}
\Delta e_f = \; & \underset{(0.035)}{0.13}\,\Delta e_{f,t-1} - \underset{(0.012)}{0.11}\,I_{31} - \underset{(0.012)}{0.11}\,I_{32} \\
& + \underset{(0.0096)}{0.028}\,I_{34} - \underset{(0.0096)}{0.027}\,I_{43} + \underset{(0.0085)}{0.031}\,I_{70} \\
& + \underset{(0.04)}{0.59}\,\Delta e_t - \underset{(0.031)}{0.32}\,\Delta(p_f - p)_t - \underset{(0.1)}{0.19}\,\Delta n_t \\
& + \underset{(0.035)}{0.23}\,\Delta s_t - \underset{(0.023)}{0.36}\,ECM_{t-1}
\end{aligned}
$$

$$
F_{ar}(2,59) = 0.68 \quad \chi^2_{nd}(2) = 1.78 \quad F_{arch}(1,70) = 0.27
$$
$$
F_{reset}(2,59) = 0.23 \quad F_{het}(12,54) = 1.01
$$

Solved cointegrating relation with dummies excluded:

$$
ECM = e_f - \underset{(0.01)}{0.63}\,e + \underset{(0.04)}{0.13}\,(p_f - p) - \underset{(0.08)}{1.12}\,s + \underset{(0.01)}{0.45}\,n
$$



Adding indicators for every observation?

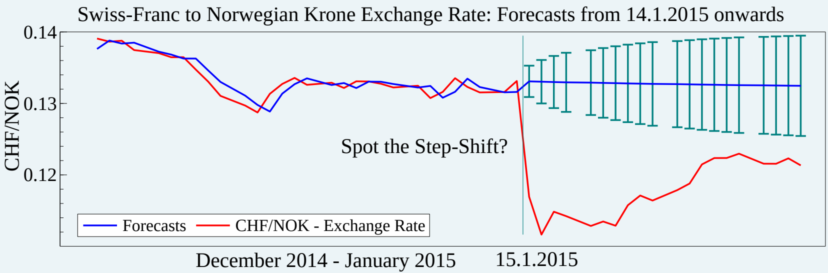
David Hendry trying to convince Søren Johansen at Engle and Granger Nobel award ceremony 2003 Stockholm:

Unknown unknowns: unknown number of location shifts/outliers of unknown magnitudes at unknown times.

- Testing model mis-specification
- Learning from data
- Testing super exogeneity

Unmodelled location shifts have pernicious effects:

- **in sample**, mis-specified empirical models, distorting inference;
- **out of sample**, assess forecast failure.



Numbers and magnitudes of breaks in models usually unknown: obviously unknown for unknowingly omitted variables.

General approach required to detect location shifts anywhere in sample **while also selecting over many candidate variables**.

Theory-embedding in general model allowing for outlier/location shift at any point in time.

Numbers and magnitudes of breaks in models usually unknown: obviously unknown for unknowingly omitted variables.

General approach required to detect location shifts anywhere in sample **while also selecting over many candidate variables**.

Theory-embedding in general model allowing for outlier/location shift at any point in time.

Impulse-Indicator Saturation (IIS) creates complete set of indicator variables: $\{1_{\{j=t\}}\} = 1$ when $j = t$ and 0 otherwise for $j = 1, \ldots, T$:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \ddots \end{bmatrix} \tag{3}$$

add T impulse indicators to set of candidate variables when T obs.

Hendry, Johansen, and Santos (2008): **impulse-indicator saturation (IIS)**: adding an indicator dummy variable for each observation to candidate set of variables for the model.

$$y_t = \alpha_0 + \alpha_1 I_1 + \alpha_2 I_2 + ... + \alpha_T I_T + u_t. \tag{4}$$

This has $T + 1$ parameters for T observations.
However, the impulses can be added in blocks (say $T = 100$):

1. Partition in 2 blocks, $B_1 = I_1, ..., I_{50}$, $B_2 = I_{51}, ..., I_{100}$, C.f. estimating models over two subsamples of $T/2$.
2. Run model selection on each block, form union S,
3. Run model selection on S.

Consider $y_t \sim$ IID $\left[\mu, \sigma_\epsilon^2\right]$ for $t = 1, \ldots, T$

- First, include half of indicators, record significant: **just 'dummying out' $T/2$ observations for estimating μ**
- Then omit, include other half, record again.
- Combine sub-sample indicators, & select significant.

αT indicators selected on average at significance level α

Feasible 'split-sample' (IIS) algorithm: see Hendry, Johansen, and Santos (2008)

Many well-known procedures are variants of IIS.

- Chow (1960) test is sub-sample IIS over $T - k + 1$ to T without selection.
- Salkever (1976) tests parameter constancy by indicators.
- Recursive estimation equivalent to IIS over future sample, reducing indicators one at a time.

Next Figure illustrates ‘split-half’ approach for $y_t \sim \mathsf{IN}\left[\mu, \sigma_y^2\right]$
Three rows correspond to the three stages:

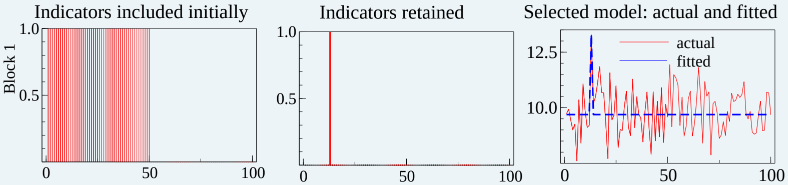
- first half of the indicators, second half, then selected indicators combined.

Three columns report:

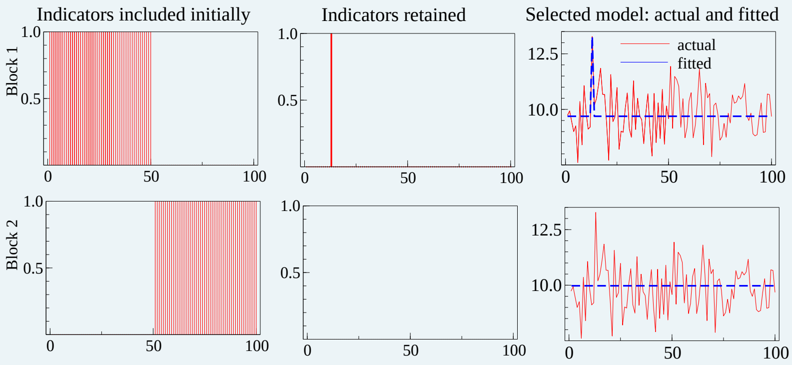
- indicators entered,
- indicators retained,
- and fitted and actual values of selected model.

Many indicators added, but only one is retained in row 1.
When second half entered (row 2), none is retained.
Combined retained dummies entered (here just one), and selection again retains it.

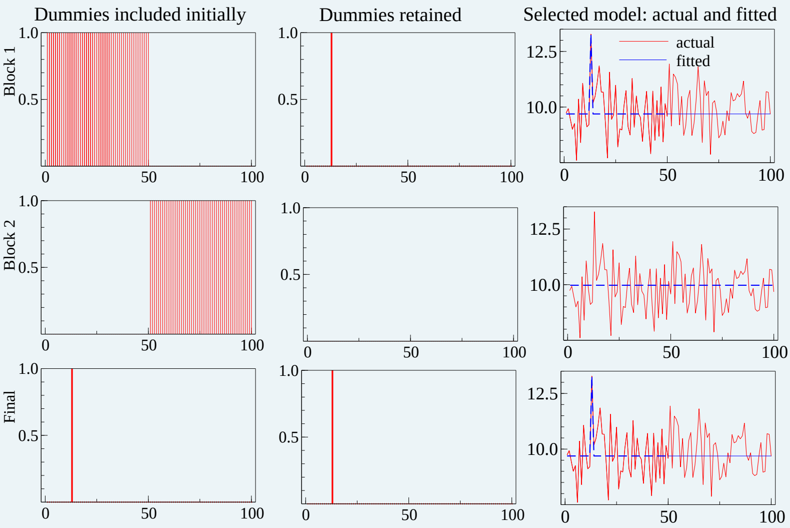
Null 'split-sample' search in IIS



Null 'split-sample' search in IIS



Null 'split-sample' search in IIS



Consider adding first half of the indicators:

$$y_t = \mu_1 + \sum_{j=1}^{T/2} \delta_j d_{j,t} + \varepsilon_t. \tag{5}$$

Consider adding first half of the indicators:

$$y_t = \mu_1 + \sum_{j=1}^{T/2} \delta_j d_{j,t} + \varepsilon_t. \tag{5}$$

The estimators are:

$$\widehat{\mu}_1 = \frac{1}{T/2} \sum_{t=T/2+1}^{T} y_t, \tag{6}$$

$$s_1^2 = \frac{1}{T/2-1} \sum_{t=T/2+1}^{T} (y_t - \widehat{\mu}_1)^2 \tag{7}$$

$$\widehat{\delta}_t = y_t - \widehat{\mu}_1, \quad t = 1, \ldots, T/2 \tag{8}$$

Consider adding first half of the indicators:

$$y_t = \mu_1 + \sum_{j=1}^{T/2} \delta_j d_{j,t} + \varepsilon_t. \tag{5}$$

The estimators are:

$$\widehat{\mu}_1 = \frac{1}{T/2} \sum_{t=T/2+1}^{T} y_t, \tag{6}$$

$$s_1^2 = \frac{1}{T/2-1} \sum_{t=T/2+1}^{T} (y_t - \widehat{\mu}_1)^2 \tag{7}$$

$$\widehat{\delta}_t = y_t - \widehat{\mu}_1, \quad t = 1, \ldots, T/2 \tag{8}$$

so that residuals are:

$$\widehat{\varepsilon}_t = 0, \quad t = 1, \ldots, T/2$$

$$\widehat{\varepsilon}_t = y_t - \widehat{\mu}_1, \quad i = T/2+1, \ldots, T$$

The final estimates are:

$$\widetilde{\mu} = \frac{\sum_{t=1}^{T_1} y_t 1_{\{|t_{1,\widehat{\delta}_t}|<c_\alpha\}} + \sum_{t=T_1+1}^{T} y_t 1_{\{|t_{2,\widehat{\delta}_t}|<c_\alpha\}}}{\sum_{t=1}^{T_1} 1_{\{|t_{1,\widehat{\delta}_t}|<c_\alpha\}} + \sum_{t=T_1+1}^{T} 1_{\{|t_{2,\widehat{\delta}_t}|<c_\alpha\}}} \tag{9}$$

and

$$\widetilde{\sigma}_\varepsilon^2 = \frac{\sum_{t=1}^{T_1} (y_t - \widehat{\mu}_1)^2 1_{\{|t_{1,\widehat{\delta}_t}|<c_\alpha\}} + \sum_{t=T_1+1}^{T} (y_t - \widehat{\mu}_2)^2 1_{\{|t_{2,\widehat{\delta}_t}|<c_\alpha\}}}{\sum_{t=1}^{T_1} 1_{\{|t_{1,\widehat{\delta}_t}|<c_\alpha\}} + \sum_{t=T_1+1}^{T} 1_{\{|t_{2,\widehat{\delta}_t}|<c_\alpha\}} - 1}. \tag{10}$$

Let $y_t = \mu + \sigma_\varepsilon \varepsilon_t$, $t = 1, \ldots, T$ be i.i.d., where ε_t has symmetric continuous density $f(.)$ with mean zero, variance one. Let $T = T_1 + T_2$ and assume that $T_1/T \to \lambda_1$ and $T_2/T \to \lambda_2$ where $0 < \lambda_1, \lambda_2 < 1$, with $\lambda_1 + \lambda_2 = 1$ then:

$$T^{1/2}\left(\widetilde{\mu} - \mu\right) \xrightarrow{D} \mathsf{N}\left[0, \sigma_\varepsilon^2 \sigma_\mu^2\right] \tag{7}$$

where

$$\sigma_\mu^2 = \left(\int_{-c_\alpha}^{c_\alpha} f(\varepsilon)d\varepsilon\right)^{-2} \left[\int_{-c_\alpha}^{c_\alpha} \varepsilon^2 f(\varepsilon)d\varepsilon(1 + 4c_\alpha f(c_\alpha)) + \left(\frac{\lambda_1^2}{\lambda_2} + \frac{\lambda_2^2}{\lambda_1}\right)(2c_\alpha f(c_\alpha))^2\right]$$

where:

$$\int_{-c_\alpha}^{c_\alpha} f(\epsilon)d\epsilon = 1 - \alpha$$

measures the impact of truncating the residuals.

Using $\int_{-c_\alpha}^{c_\alpha} f(\varepsilon)d\varepsilon = 1 - \alpha$, and for the normal distribution, $f(\varepsilon) = \phi(\varepsilon)$, using integration by parts we find:

$$\int_{-c_\alpha}^{c_\alpha} \varepsilon^2 \phi(\varepsilon)d\varepsilon = \int_{-c_\alpha}^{c_\alpha} \phi(\varepsilon)d\varepsilon - 2c_\alpha\phi(c_\alpha),$$

so that for $\lambda_1 = \lambda_2 = 0.5$ (split-half) above simplifies to:

$$\sigma_\mu^2 = \frac{1}{(1-\alpha)}\left(1 + 4c_\alpha\phi(c_\alpha) - \frac{2c_\alpha\phi(c_\alpha)}{(1-\alpha)}\left[1 + 2c_\alpha\phi(c_\alpha)\right]\right)$$

where $\sigma_\mu^2 \to 1$ as $|c_\alpha| \to \infty$.

$$\widetilde{\sigma}_{\varepsilon}^{2} = \frac{\sum_{i=1}^{T_1}(y_t - \widehat{\mu}_1)^2 1_{\{|t_{1,\widehat{\delta}_t}|<c_\alpha\}} + \sum_{t=T_1+1}^{T}(y_t - \widehat{\mu}_2)^2 1_{\{|t_{2,\widehat{\delta}_t}|<c_\alpha\}}}{\sum_{t=1}^{T_1} 1_{\{|t_{1,\widehat{\delta}_t}|<c_\alpha\}} + \sum_{t=T_1+1}^{T} 1_{\{|t_{2,\widehat{\delta}_t}|<c_\alpha\}} - 1}. \tag{11}$$

The estimator $\widetilde{\sigma}_{\varepsilon}^{2}$, has the limit

$$\widetilde{\sigma}_{\varepsilon}^{2} \xrightarrow{P} \sigma_{\varepsilon}^{2}\kappa = V(\varepsilon||\varepsilon| < c_\alpha).$$

For the normal distribution, $f(\varepsilon) = \phi(\varepsilon)$,we have the expression:

$$\kappa = 1 - \frac{2c_\alpha \phi(c_\alpha)}{1-\alpha}.$$

where $\kappa \to 1$ as $|c| \to \infty$

Effects of Impulse Indicator Saturation under null:

Selection effects on mean and variance estimators similar to 'trimming':

- Small loss in efficiency (consistency effect on variance estimate $\tilde{\sigma}_{\epsilon}$, efficiency effect through σ^2_{μ})
- Controllable by choosing more conservative α

IIS interpretable as robust estimator, but allows for joint selection over variables.

Johansen and Nielsen (2016) 'gauge' is consistent:

$$\hat{g} = \frac{1}{T}\sum_{t=1}^{T} 1_{(|y_t - x_t\tilde{\beta}| > \tilde{\sigma}_{\epsilon} c_{\alpha})} \tag{12}$$

$$E[\hat{g}] \rightarrow P(|\epsilon_1| > \sigma c_{\alpha}) = \alpha \tag{13}$$

Johansen & Nielsen (2009) & (2016) extend IIS to both stationary and unit-root autoregressions:

When distribution is symmetric, adding T impulse-indicators to a regression with n variables, coefficient $\boldsymbol{\beta}$ (not selected) and second moment $\boldsymbol{\Sigma}$:

$$T^{1/2}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} \mathsf{N}_n \left[\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{\Sigma}^{-1} \boldsymbol{\Omega}_\beta\right]$$

Efficiency of IIS estimator $\widetilde{\boldsymbol{\beta}}$ with respect to OLS $\widehat{\boldsymbol{\beta}}$ measured by $\boldsymbol{\Omega}_\beta$ depends on c_α and distribution

Johansen & Nielsen (2009) & (2016) extend IIS to both stationary and unit-root autoregressions:

When distribution is symmetric, adding T impulse-indicators to a regression with n variables, coefficient $\boldsymbol{\beta}$ (not selected) and second moment $\boldsymbol{\Sigma}$:

$$T^{1/2}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} \mathsf{N}_n \left[\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{\Sigma}^{-1} \boldsymbol{\Omega}_\beta\right]$$

Efficiency of IIS estimator $\widetilde{\boldsymbol{\beta}}$ with respect to OLS $\widehat{\boldsymbol{\beta}}$ measured by $\boldsymbol{\Omega}_\beta$ depends on c_α and distribution

Must lose efficiency under null; small loss αT: 1 observation at $\alpha = 1/T$ if $T = 100$, despite T extra candidates.

Potential for major gain under alternatives of breaks and/or data contamination: variant of robust estimation
but can be done jointly with all other selections

Single Outlier at $t = T_1$:

DGP: $y_t = \lambda_1 1_{\{t=T_1\}} + \epsilon_t$

matched by **model**: $y_t = \gamma d_{t=T_1}$

$$
\begin{aligned}
(\hat{\gamma} - \lambda_1) &= (d'_{t=T_1} d_{t=T_1})^{-1} d'_{t=T_1} \epsilon_t \\
&= \epsilon_{T_1}
\end{aligned}
$$

Unbiased but not consistent (Hendry and Santos (2005)). With variance:

$$V[\hat{\gamma}] = V[\epsilon_{T_1}] = \sigma_\epsilon^2$$

t-statistic:

$$t_{\hat{\gamma}} = \frac{\hat{\gamma}}{\hat{\sigma}_\epsilon} \approx \frac{(\hat{\gamma} - \lambda_1)}{\sigma_\epsilon} + \frac{\lambda_1}{\sigma_\epsilon} \sim N(\psi_{\lambda_1}, 1)$$

Illustrate IIS for a location shift of λ over last k observations:

$$y_t = \mu + \lambda 1_{\{t \geqslant T-k+1\}} + \varepsilon_t \tag{14}$$

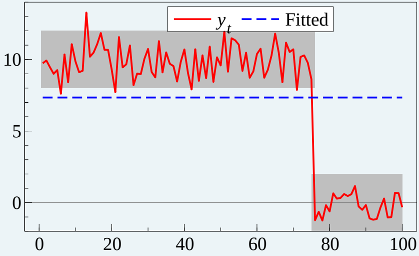
where $\varepsilon_t \sim \mathsf{IN}\left[0, \sigma_\varepsilon^2\right]$ and $\lambda \neq 0$.

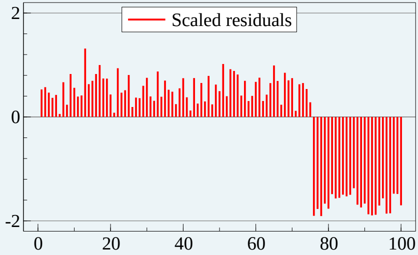
Optimal test is t-test for a break in (14) at $T - k + 1$ onwards, requires:

- knowledge of location-shift timing
- knowing that it is the only break
- is same magnitude break thereafter

The next slide records IIS for $\lambda = 10\sigma_\varepsilon$ in (14) at $0.75T = 75$.

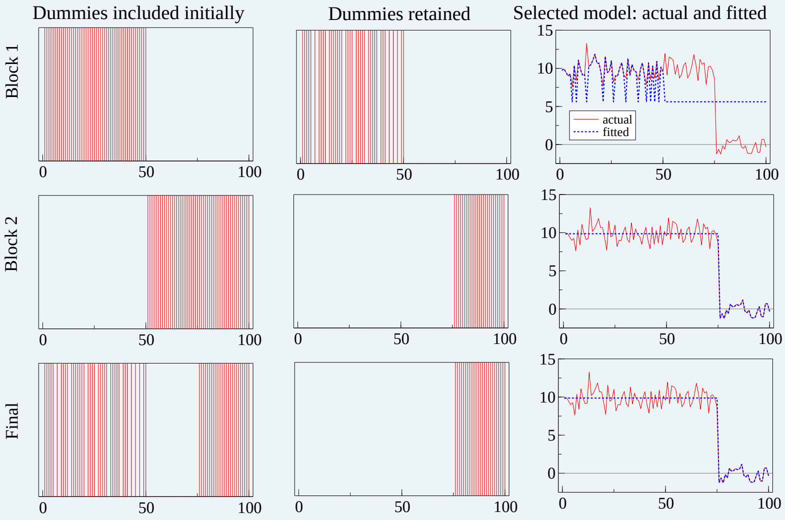
Structural break example





- Size of the break is **10 standard errors** at $0.75T$
- There are **no outliers** in this mis-specified model as all residuals $\in [-2, 2]$ SDs:

  outliers $\neq$ structural breaks
- step-wise regression has **zero power**


Dummies included initially
Dummies retained
Selected model: actual and fitted
Block 1
Block 2
Final
actual
fitted
0
50
100
0
5
10
15

- Initially, many indicators now retained (top row), considerable discrepancy between the first-half and second-half means.
- When second set entered, all indicators for location shift period are retained.
- Once combined set entered, despite large number of dummies, selection reverts to just those for break period.

Under null, indicators significant in sub-sample would remain so overall, for alternatives, sub-sample significance can be transient, due to unmodeled features that occur elsewhere in data.

Extension of IIS to step-indicator saturation (SIS):
Regression model saturated with complete set of step indicators

$$\mathcal{S}_1 = \left\{1_{\{t \leqslant j\}}, j = 1, \ldots, T\right\}$$

where $1_{\{t \leqslant j\}} = 1$ for observations up to j, and zero otherwise.

Step indicators cumulate impulse indicators up to each next observation:

IIS: Impulses SIS: Step shifts

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \ddots \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

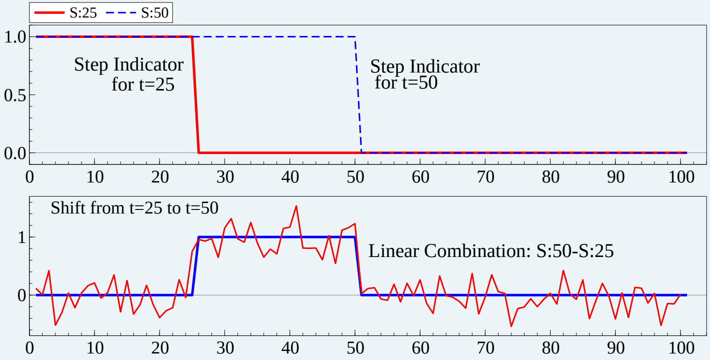
Step indicators take the form:
$\boldsymbol{\iota}_1' = (1, 0, 0, \ldots, 0),\ \boldsymbol{\iota}_2' = (1, 1, 0, \ldots, 0),\ \ldots, \boldsymbol{\iota}_T' = (1, 1, 1, \ldots, 1),$

Extension of IIS to step-indicator saturation (SIS):

Regression model saturated with complete set of step indicators

$$\mathcal{S}_1 = \{1_{\{t \leqslant j\}}, j = 1, \ldots, T\}$$

where $1_{\{t \leqslant j\}} = 1$ for observations up to j, and zero otherwise.



IIS and SIS - important differences necessitate a new analysis:

- Impulse Indicators: mutually orthogonal. **Step indicators overlap** increasingly as their second index increases.
- Two indicators are required to characterize an outlier or shift not at the end of the sample: $1_{\{t \leqslant T_2\}} - 1_{\{t < T_1\}}$.
- Opens the door to "**designed break functions**" (Volcanoes! See Pretis, Schneider, Smerdon, and Hendry (2016))

The Step-Indicator Saturation Model (Infeasible as $N >> T$):

$$y_t = \beta_0 + \boldsymbol{\beta}_1' \mathbf{z}_t + \sum_{j=1}^{T-1} \delta_j 1_{\{t \leqslant j\}} + \epsilon_t \quad \text{where} \quad \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{15}$$

Split-Half Approach:

- Add the first $T/2$ indicators from the saturating set $\mathcal{S}_1$:

$$y_t = \beta_0 + \boldsymbol{\beta}_1' \mathbf{z}_t + \sum_{j=1}^{T/2} \delta_j 1_{\{t \leqslant j\}} + \epsilon_t \tag{16}$$

  can be estimated directly, indicators retained when estimated coefficients $\widehat{\delta}_j$ satisfy $\left|t_{\widehat{\delta}_j}\right| > c_\alpha$ where c_α is the critical value for significance level α.
- Locations are recorded, all those indicators are dropped, second set is then investigated.
- Combine selected indicators and re-select.

- Under the null with $\alpha = 1/T$, at both sub-steps on average, $\alpha T/2$ (namely $1/2$ an indicator) will be retained by chance.
- On average $\alpha T = 1$ indicator will be retained from the combined stage: **gauge should equal nominal size.**
- One degree of freedom is lost on average.

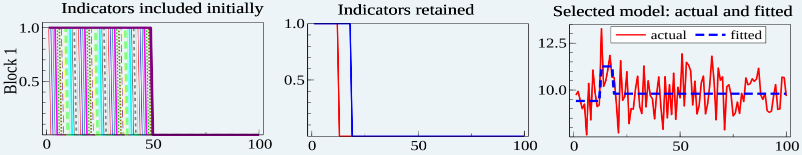
When m indicators are selected in a congruent representation at significance level α:

$$y_t = \beta_0 + \boldsymbol{\beta}_1' \mathbf{z}_t + \sum_{i=1}^{m} \phi_{i,\alpha} 1_{\{t \leqslant T_i\}} + \nu_t \tag{17}$$

where $\nu_t \sim \mathsf{IN}\left[0, \sigma_\nu^2\right]$, and coefficients of significant indicators are denoted $\phi_{i,\alpha}$.

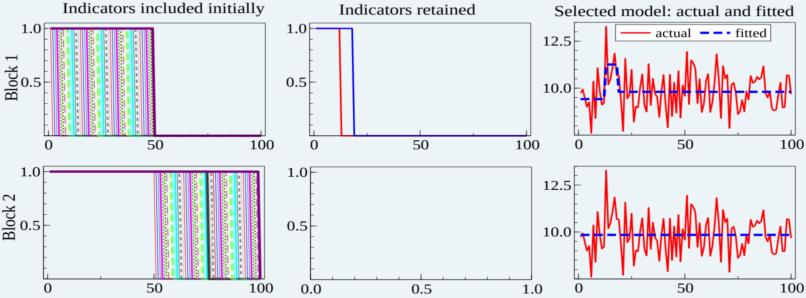
OXFORD MARTIN SCHOOL UNIVERSITY OF OXFORD Institute for New Economic Thinking AT THE OXFORD MARTIN SCHOOL

‘Split-sample’ search by SIS at 1%.

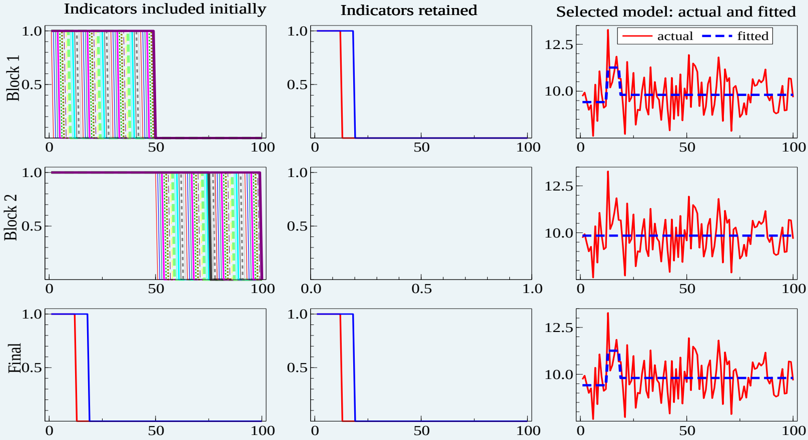


OXFORD MARTIN SCHOOL UNIVERSITY OF OXFORD Institute for New Economic Thinking AT THE OXFORD MARTIN SCHOOL

‘Split-sample’ search by SIS at 1%.



Illustrating SIS when no location shifts

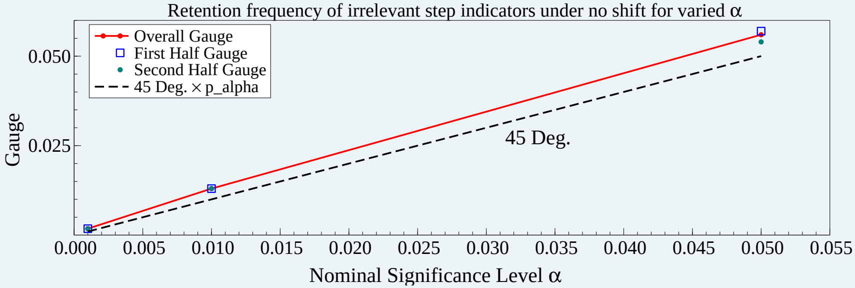


$T = 100$, **and no shifts, retains 2 significant steps, so lose 2 degrees of freedom–but could be combined to one dummy.**

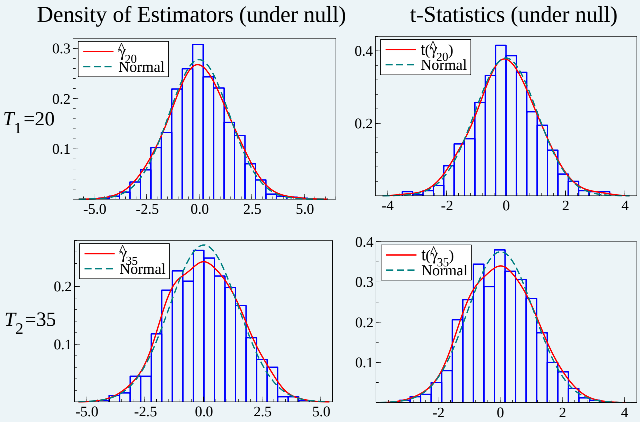
Under the null of no shift:

Model with $n = 0$: using the split-half approach where $\beta_0 = 0$, $\epsilon_t \sim \mathsf{IN}[0, \sigma^2_\epsilon]$ and $\sigma^2_\epsilon = 1$, for a sample size $T = 100$ and various values of α.

Retention frequency of irrelevant indicators: close to α, on average αT irrelevant step indicators retained under the null:



Properties of two out of the 100 step indicators ($T_1 = 20, T_2 = 35$) under the null of no shift:



Both estimators $\widehat{\gamma}_{T_1}$ and $\widehat{\gamma}_{T_2}$ have densities close to Normal, centered on zero, and central t-statistics.

Known mean shift from $\lambda_1 \neq 0$ to $\lambda_1 = 0$ at time $0 < T_1 < T/2$ in DGP:

$$y_t = \mu + \lambda_1 1_{\{t \leqslant T_1\}} + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}\left[0, \sigma^2_\epsilon\right] \tag{18}$$

where $\lambda_1 \neq 0$: shift is from $\mu + \lambda_1$ to μ.
Nesting model of (18) when the break is known:

$$y_t = \varphi + \delta_{T_1} 1_{\{t \leqslant T_1\}} + \nu_t \tag{19}$$

As $\sum_{t=1}^{T} 1_{\{t \leqslant T_1\}} = \sum_{t=1}^{T_1} 1_{\{t \leqslant T_1\}} = T_1$, estimating (19) delivers:

$$\begin{pmatrix} \widehat{\varphi} - \mu \\ \widehat{\delta}_{T_1} - \lambda_1 \end{pmatrix} = \begin{pmatrix} T & T_1 \\ T_1 & T_1 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T} \epsilon_t \\ \sum_{t=1}^{T_1} \epsilon_t \end{pmatrix} = \begin{pmatrix} \overline{\epsilon}_{(2)} \\ \overline{\epsilon}_{(1)} - \overline{\epsilon}_{(2)} \end{pmatrix}$$

where $\overline{\epsilon}_{(1)} = T_1^{-1} \sum_{t=1}^{T_1} \epsilon_t$ average over first T_1 observations and $\overline{\epsilon}_{(2)} = (T - T_1)^{-1} \sum_{t=T_1+1}^{T} \epsilon_t$

And the variance is:

$$V\left[\begin{pmatrix} \widehat{\varphi}-\mu \\ \widehat{\delta}_{T_1}-\lambda_1 \end{pmatrix}\right] = \sigma^2_\epsilon (T-T_1)^{-1} \begin{pmatrix} 1 & -1 \\ -1 & T_1^{-1}(T-T_1)+1 \end{pmatrix}.$$

For the DGP in (18):

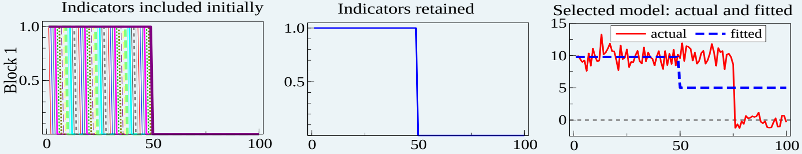
$$\sqrt{T^*}\left(\widehat{\delta}_{T_1}-\lambda_1\right) \sim \mathsf{N}\left[0, \sigma^2_\epsilon\right] \tag{20}$$

Hence, neglecting the estimation uncertainty in $\widehat{\sigma}^2_\epsilon$ and letting $(T_1^{-1}+(T-T_1)^{-1})^{-1}=T^*$:

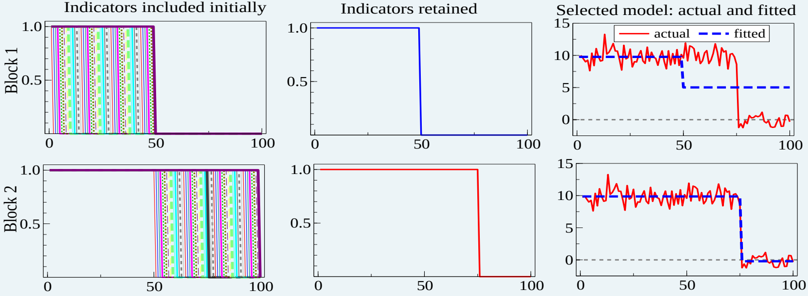
$$t_{\widehat{\delta}_{T_1}} = \frac{\sqrt{T^*}\widehat{\delta}_{T_1}}{\widehat{\sigma}_\epsilon} \approx \frac{\sqrt{T^*}(\widehat{\delta}_{T_1}-\lambda_1)}{\sigma_\epsilon} + \frac{\sqrt{T^*}\lambda_1}{\sigma_\epsilon} \sim \mathsf{N}\left[\psi^*_{\lambda_1}, 1\right] \tag{21}$$

where $T^*=T_1$ when there is no intercept. Yields $\sqrt{T^*}$ times the corresponding non-centrality for an individual impulse indicator.

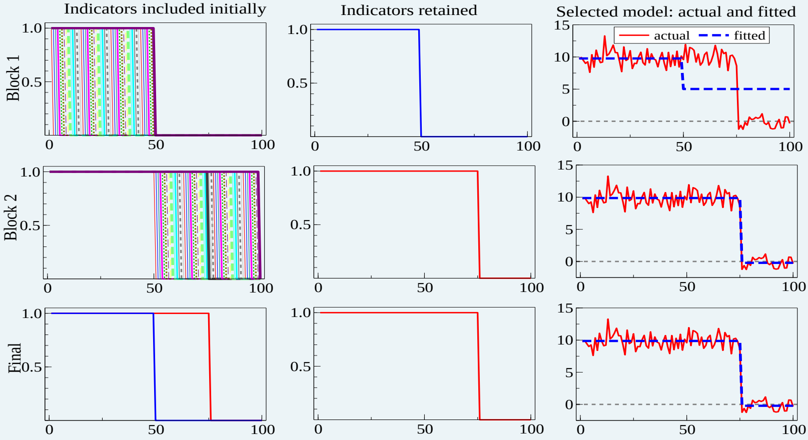
Add half indicators and select ones significant at 1%.



Drop, add other half indicators and again select at 1%.



Combine retained indicators and re-select at 1%.



Matching theory: initially retains last step indicator closest to mean shift, then finds correct shift, so eliminates redundant indicator. Just one step indicator needed.

Detection of single location shift falling within a half-sample of the data ($0 < T_1 < T/2$) using split-half analysis of SIS where DGP is:

$$\mathbf{y} = \lambda_1 \boldsymbol{\iota}_{T_1} + \boldsymbol{\epsilon} \tag{22}$$

Add first half of step indicators, model is:

$$y_t = \sum_{j=1}^{T/2} \gamma_j 1_{\{t \leqslant j\}} + \nu_t \tag{23}$$

Intercept of zero highlights main aspects of the algebra, written as:

$$\mathbf{y} = \mathbf{D}_1 \boldsymbol{\gamma}_{(1)} + \mathbf{v} \tag{24}$$

where $\boldsymbol{\gamma}_{(1)} = (\gamma_1 \dots \gamma_{T/2})'$ and $\mathbf{D}_1 = (\boldsymbol{\iota}_1 \dots \boldsymbol{\iota}_{T/2})$. Then:

$$\widehat{\gamma}_{(1)} = \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1} \mathbf{D}_1'\mathbf{y} = \lambda \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1} \mathbf{D}_1'\boldsymbol{\iota}_{T_1} + \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1} \mathbf{D}_1'\boldsymbol{\epsilon} \tag{25}$$

The inverse of $\left(\mathbf{D}_1'\mathbf{D}_1\right)$ is the 'double difference' matrix:

$$\left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1} = \begin{pmatrix} 2 & -1 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 \\ 0 & -1 & 2 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix} \tag{26}$$

so:

$$\left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\mathbf{D}_1' = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

is the forward-difference matrix.

Letting $\nabla\epsilon_t = \epsilon_t - \epsilon_{t+1}$, from $\widehat{\gamma}_{(1)} = \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\mathbf{D}_1'\mathbf{y}$:

$$\widehat{\gamma}_{(1)} = \lambda_1\mathbf{r} + \nabla\boldsymbol{\epsilon}_{(1)}$$

where $\mathbf{r}$ is a $T/2 \times 1$ vector with unity at $t = T_1$and zeroes elsewhere, so:

$$\left(\widehat{\gamma}_{(1)} - \lambda_1\mathbf{r}\right) = \nabla\boldsymbol{\epsilon}_{(1)} \tag{27}$$

where the $(T/2 \times 1)$ vector $\nabla\boldsymbol{\epsilon}_{(1)} = (\nabla\epsilon_1, \nabla\epsilon_2, \ldots, \nabla\epsilon_{T/2}, \epsilon_{T/2})'$. All elements of $\widehat{\gamma}_{(1)}$ up to the T_1th are zero, only the T_1th reflect λ_1, corresponding to the location shift.

Only the value of λ_1 at the shift is being picked up, incremental information equivalent to an impulse indicator for T_1:

$$\widehat{\gamma}_{T_1} = \lambda_1 + \nabla\epsilon_{T_1} \tag{28}$$

Hence:

$$\left(\widehat{\gamma}_{(1)} - \lambda_1 \mathbf{r}\right)\underset{\widetilde{app}}{} \mathsf{N}\left[\mathbf{0}, \sigma_\epsilon^2 \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\right] \tag{29}$$

The estimated error variance adjusted for degrees of freedom:

$$\widehat{\sigma}_\epsilon^2 = \frac{2}{T}\sum_{t=T/2+1}^{T}\left(y_t - \widehat{y}_t\right)^2$$

will be an unbiased estimator of σ_ϵ^2. However, for IID errors (because of $\nabla \epsilon_{T_1}$):

$$\mathsf{V}\left[\widehat{\gamma}_{T_1}\right] = 2\sigma_\epsilon^2 \tag{30}$$

so that:

$$t_{\widehat{\gamma}_{T_1}} = \frac{\widehat{\gamma}_{T_1}}{\sqrt{2}\widehat{\sigma}_\epsilon} \approx \frac{\left(\widehat{\gamma}_{T_1} - \lambda_1\right)}{\sqrt{2}\sigma_\epsilon} + \frac{\lambda_1}{\sqrt{2}\sigma_\epsilon} \sim \mathsf{N}\left[\frac{\psi_{\lambda_1}}{\sqrt{2}}, 1\right] \tag{31}$$

where $\psi_{\lambda_1}/\sqrt{2}$ is the non-centrality. Note: indep. of length of shift.

Selection is essential

$\mathsf{V}\left[\widehat{\gamma}_{T_1}\right] = 2\sigma_\epsilon^2$ due to collinearity between step indicators:

Eliminating insignificant indicators by sequential selection or multi-path search is essential.

Example: At 1%, $c_\alpha \approx 2.7$, normalizing on $\sigma_\epsilon = 1$, requires $\lambda_1 > 3.8$ for even a 50% chance of significance before simplification.

When insignificant indicators are deleted, $\mathsf{V}[\widehat{\gamma}_{T_1}]$ falls rapidly:

If all irrelevant indicators eliminated, just $\boldsymbol{\iota}_{T_1}$ remains, the non-centrality for a single shift $\psi_1 = \sqrt{T^*}\lambda_1/\sigma_\epsilon$ which is $\sqrt{2T^*}$ larger than before selection.

First half step indicators are then eliminated and second half, $\mathbf{D}_2 = (\boldsymbol{\iota}_{T/2+1} \ldots \boldsymbol{\iota}_T)$ added.

- If significant steps from first half retained: only $\alpha/2$ of estimated coefficients of $\mathbf{D}_2$ should be significant
- If significant steps are not retained then model becomes:

$$y_t = \sum_{j=T/2+1}^{T} \gamma_j 1_{\{t \leqslant j\}} + \nu_t \tag{32}$$

written as:

$$\mathbf{y} = \mathbf{D}_2 \boldsymbol{\gamma}_{(2)} + \mathbf{v} \tag{33}$$

where $\boldsymbol{\gamma}_{(2)} = (\gamma_{T/2+1} \ldots \gamma_T)'$ and $\mathbf{D}_2 = (\boldsymbol{\iota}_{T/2+1} \ldots \boldsymbol{\iota}_T)$. From (22):

$$\widehat{\gamma}_{(2)} = \left(\mathbf{D}_2'\mathbf{D}_2\right)^{-1} \mathbf{D}_2'\mathbf{y} = \lambda_1 \left(\mathbf{D}_2'\mathbf{D}_2\right)^{-1} \mathbf{D}_2' \boldsymbol{\iota}_{T_1} + \left(\mathbf{D}_2'\mathbf{D}_2\right)^{-1} \mathbf{D}_2' \boldsymbol{\epsilon} \tag{34}$$

$$\left(\mathbf{D}_2'\mathbf{D}_2\right) = \left(\mathbf{D}_1'\mathbf{D}_1\right) + \frac{1}{2}\mathsf{T}\mathbf{c}\mathbf{c}'$$

where $\mathbf{c}$ is a $\mathsf{T}/2 \times 1$ vector of ones and $\mathbf{j}$ is a $\mathsf{T}/2 \times 1$ vector of zeroes other than unity at first element:

$$\left(\mathbf{D}_2'\mathbf{D}_2\right)^{-1} = \left(\mathbf{I}_{\mathsf{T}/2} + \frac{\mathsf{T}}{2}\mathbf{j}\mathbf{c}'\right)^{-1}\left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}$$

and:

$$\widehat{\gamma}_{(2)} = \lambda_1 \mathsf{T}_1 \left(\mathbf{I}_{\mathsf{T}/2} + \frac{\mathsf{T}}{2}\mathbf{j}\mathbf{c}'\right)^{-1}\mathbf{j} + \left(\mathbf{D}_2'\mathbf{D}_2\right)^{-1}\mathbf{D}_2'\epsilon \qquad (35)$$

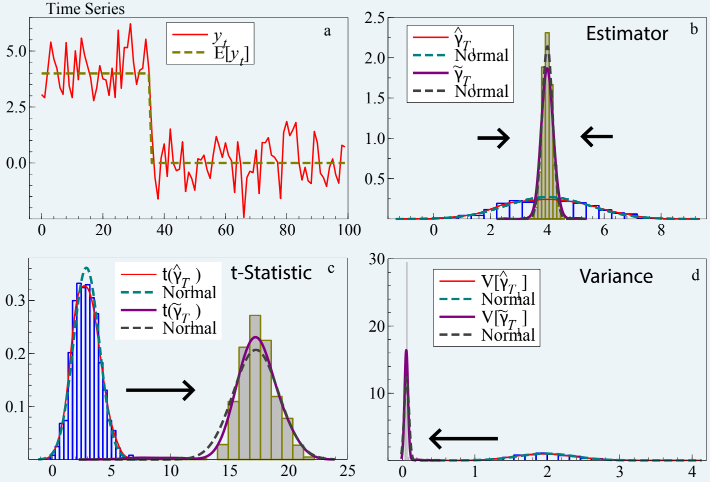
- Only first element of $\widehat{\gamma}_{(2)}$ depends on λ_1: indicator nearest to shift retained if relevant indicators not ‘carried forward’.

Finally, combine selected step indicators and re-select. When all irrelevant indicators are removed and the relevant one retained:

$$y_t = \gamma_{T_1} 1_{\{t \leqslant T_1\}} + \nu_t \tag{36}$$

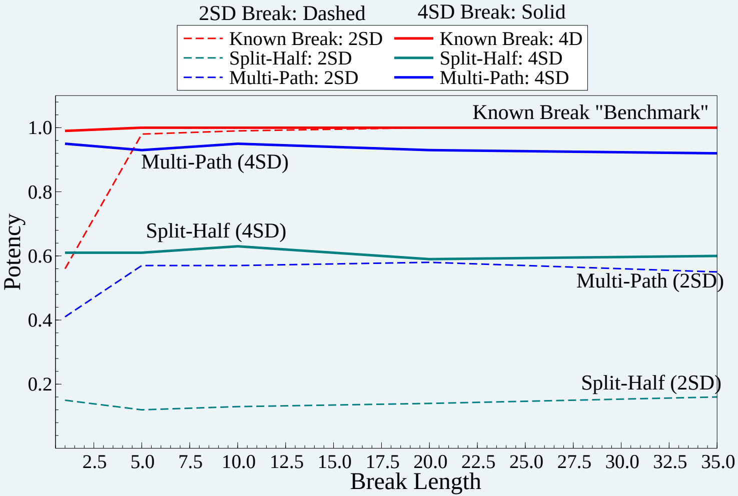
perfect selection coincides with DGP; retained irrelevant indicators reduce degrees of freedom, and increase variances.

Location Shift at $T_1 = 35$, magnitude $\lambda_1 = 4\sigma_\epsilon$, selection at $\alpha = 0.01$.



Sequential selection (grey) reduces variance vs. split-half (open, blue).

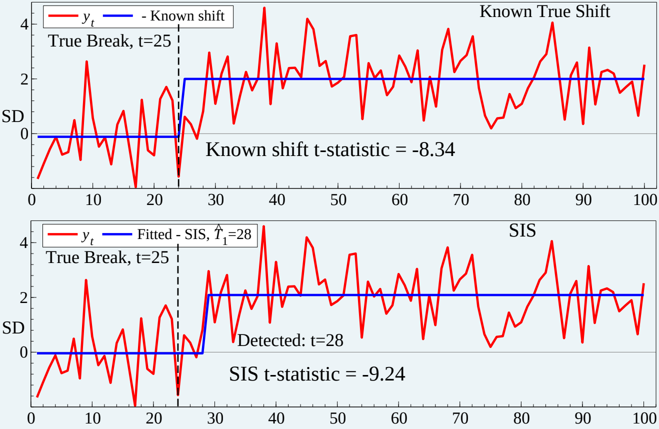
Exact ($\hat{T}_1 = T_1$) retention frequency of break for alternative methods, varying break lengths l and magnitudes, λ_1



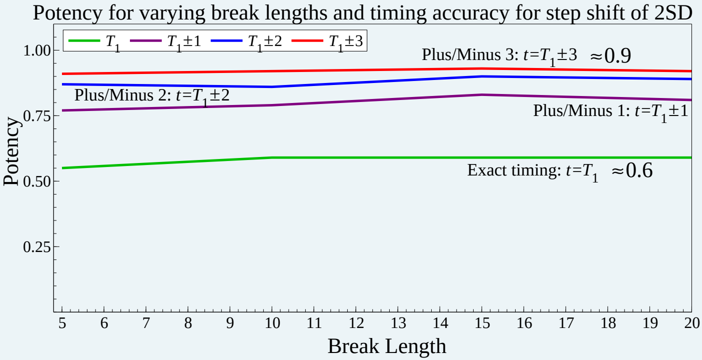
Mis-selected indicator

Selected step indicators may not exactly match location shift

Random draws of error: Mis-timed break indicator for shift at $t = 25$:



SIS selection can 'miss' by periods. **Low potency primarily due to mis-timing rather than not detecting the shift**:

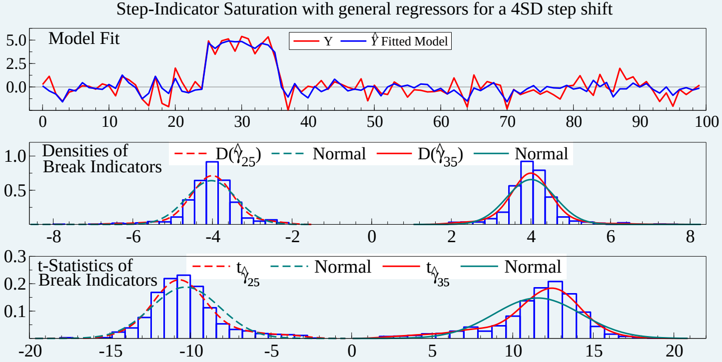


Even for $\lambda = 2\sigma_\epsilon$ and short breaks, potency is 0.9 or higher by $T_1 \pm 3$.

Simulate SIS with $n < T$ general regressors, with single step shift with unknown timing requiring two indicators, the DGP is:

$$y_t = \beta_1' \mathbf{z}_t + \lambda_1 \left(1_{\{t \leqslant T_2\}} - 1_{\{t \leqslant T_1\}}\right) + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \quad (37)$$

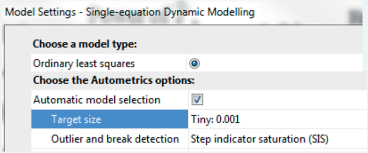
Even including 10 relevant regressors (not selected over), densities of the two shift estimators, $\widehat{\gamma}_i$, centered around true value $\lambda_1 = 4\sigma_\epsilon$:



Potency unaffected by regressors (≈ 0.5 for 2SD, ≈ 0.9 for 4SD)

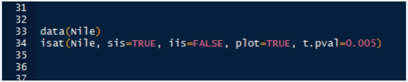
- Step and Impulse-Indicator Saturation (SIS, IIS) in *PcGive/Ox*
- `isat` in in *R*-package ‘`gets`’ (Pretis et al. 2016) (with Genaro Sucarrat and James Reade)
- Note: SIS construction differs between PcGive & R
  - PcGive: $d_{T_1} = 1_{\{t \leqslant T_1\}}$, $\hat{\gamma} > 0$ implies negative shift.
  - R: $d_{T_1} = 1_{\{t \geqslant T_1\}}$, $\hat{\gamma} > 0$ implies positive shift.

SIS in PcGive/Ox:




```
model.Autometrics(0.001, “SIS”, ...);
```

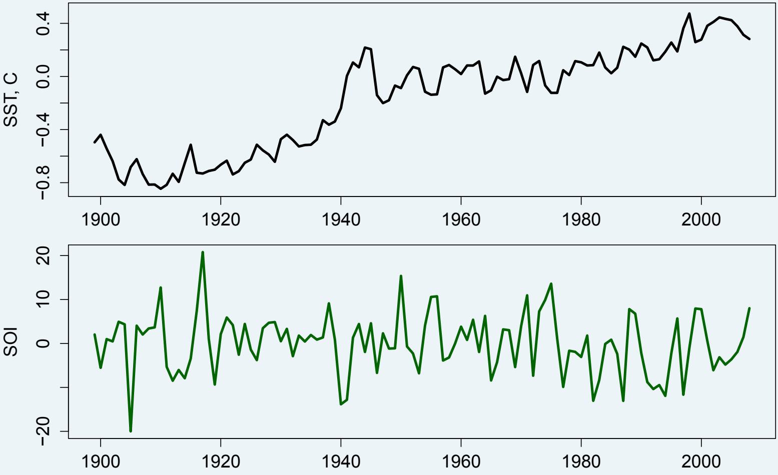
SIS in R: `isat` (gets)




```
isat(y, mxreg=..., ar=1:2,
     sis=TRUE, iis=FALSE, t.pval=0.005,...)
```

Global SST – Climate/Weather Indicator

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



Model:

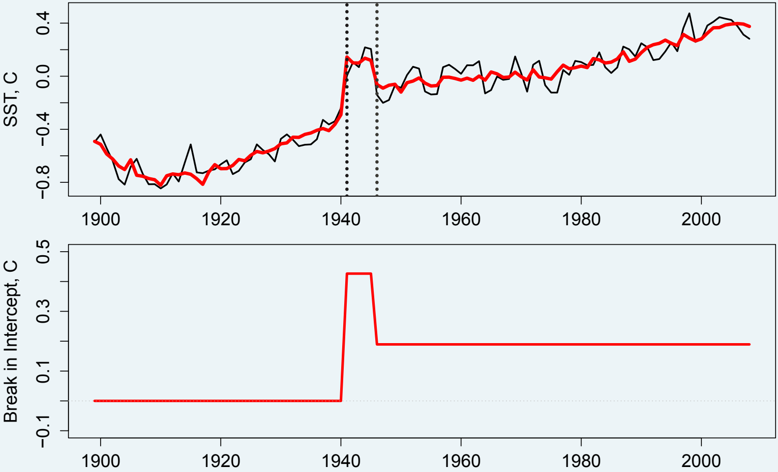
$$\mathbf{y_t} = \mathbf{f}(\mathbf{z_t}) + \beta \mathbf{x_t} + \delta \mathbf{D} + \epsilon_\mathbf{t} \qquad (38)$$

- **T**=108 (1900-2008)
- **y**= Global Mean SST Anomalies (C) relative to 1950-79
- **z**=t, **x**= Southern Oscillation Index (SOI) (atm. pressure at SL)
- **N** = 108 + 6 = 114 variables

Specification:

- SIS, $p_\alpha = 0.001$ (0.1%)
- Nonlinear trend: B-spline basis (5 degree polynomial)

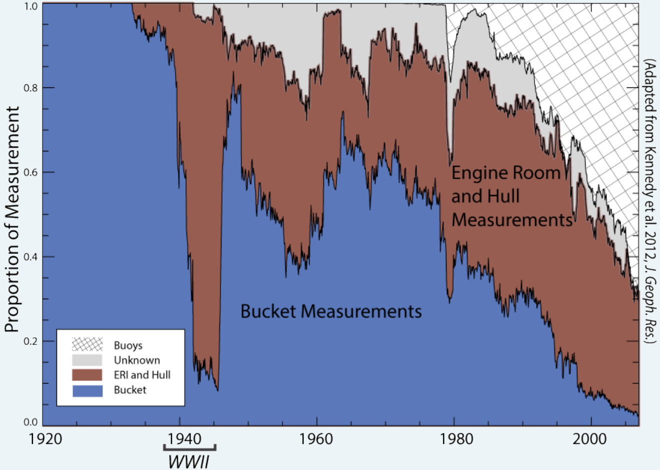
SST: SIS Results (Fitted)



Two breaks: **1941**, $\hat{\delta}_1 = 0.43^{***}\mathrm{C}$ (se=0.046)

1946, $\hat{\delta}_2 = -0.24^{***}\mathrm{C}$ (se=0.048)

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



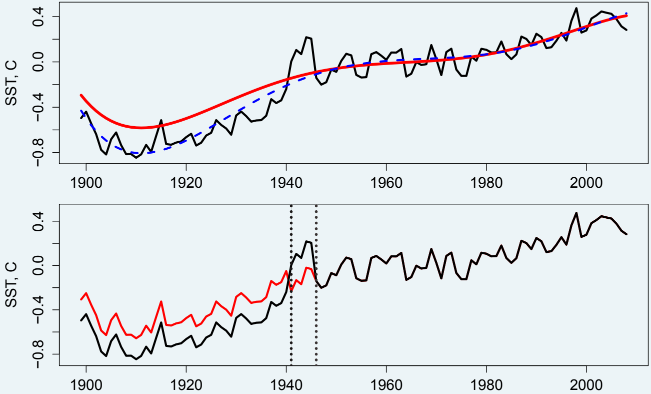
- WWII: **1941/1942** **Measurements**: buckets to engine intake
  - Danger of measurements (light)
  - Americans joined 1941/1942
- Post-WWII: Partly changed back
- Buckets: **cold bias** ($\approx 0.3C$) (Matthews, 2012)
  - SIS: $\hat{\delta}_1 = 0.43$ (± 0.046)

SOI Effect (linear)

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

Breaks

- Correct SST record for '**bucket bias**'



Commence with general formulation – **general unrestricted model**:

$$y_t = \beta' \mathbf{z}_t + \gamma' \mathbf{w}_t + \sum_{j=1}^{T} \delta_{\mathrm{IIS},j} \mathbf{1}_{\{j=t\}} + \sum_{j=1}^{T-1} \delta_{\mathrm{SIS},j} \mathbf{1}_{\{j \leqslant t\}} + \nu_t \quad t = 1, \ldots, T$$

- Embed theory z_t
- Expand model w_t (almost costless if theory correct)
- Indicators δ_t (almost costless under null)
- Ensuring valid conditioning – exogeneity

References I

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

Chow, G. C. (1960).
Tests of equality between sets of coefficients in two linear regressions.
Econometrica 28, 591–605.

Hendry, D. F. (1999).
An econometric analysis of US food expenditure, 1931–1989.
See Magnus and Morgan (1999), pp. 341–361.

Hendry, D. F., S. Johansen, and C. Santos (2008).
Automatic selection of indicators in a fully saturated regression.
Computational Statistics 23, 337–339.

Hendry, D. F. and G. E. Mizon (2011).
Econometric modelling of time series with outlying observations.
Journal of Time Series Econometrics 3(1).

Hendry, D. F. and C. Santos (2005).
Regression models with data-based indicator variables.
Oxford Bulletin of Economics and statistics 67(5), 571–595.

Johansen, S. and B. Nielsen (2009).
An analysis of the indicator saturation estimator as a robust regression estimator.
pp. 1–36. Oxford: Oxford University Press.

References II

Johansen, S. and B. Nielsen (2016).
Asymptotic theory of outlier detection algorithms for linear time series regression models.
Scandinavian Journal of Statistics 43(**2**), **321–348.**

Magnus, J. R. and M. S. Morgan (Eds.) (1999).
Methodology and Tacit Knowledge: Two Experiments in Econometrics.
Chichester: John Wiley and Sons.

Pretis, F., L. Schneider, J. E. Smerdon, and D. F. Hendry (2016).
Detecting volcanic eruptions in temperature reconstructions by designed break-indicator saturation.
Journal of Economic Surveys, **10.1111/joes.12148.**

Reade, J. J. (2008).
Updating Tobin's food expenditure time series data.
Working paper, Department of Economics, University of Oxford.

Salkever, D. S. (1976).
The use of dummy variables to compute predictions, prediction errors and confidence intervals.
Journal of Econometrics 4, **393–397.**

Tobin, J. (1950).
A statistical demand function for food in the U.S.A.
A, 113(**2**), **113–141.**