

Forecasting, Structural Breaks and Non-linearities

David F. Hendry

with

**Jennifer L. Castle,
Nicholas Fawcett and James Reade**

Department of Economics, Oxford University
ESRC Funded under Grant RES-062-23-0061

Introduction

Objectives:

develop methods for forecasting breaks; with robust strategies if breaks incorrectly predicted.

First requires that:

- (1) breaks are predictable;
- (2) we have information relevant to that predictability;
- (3) that information is available at the forecast origin;
- (4) we have a forecasting model that embodies it;
- (5) we have a method for selecting that model.

Second builds on considerable recent research:

- (6) robust forecasting devices;
- (7) improved intercept corrections;
- (8) pooling of forecasts.

Research strategy

(1) **Unpredictability and role of information:** analyzed in **Clements and Hendry (2005)**.

Suggests new formulation with **two information sets**, **which potentially might be very different**—**one economics**: regular forces from agents' behaviour; **other could be politics** (say): causes of sudden shifts.

(2) **Forecast-error taxonomy: which breaks matter?**
Location shifts are most pernicious:
induce non-stationarity & systematic forecast failure.
Theory in Clements and Hendry (1998, 2006)

(3) **'Leading indicators'**, and rapid information updates at forecast origin—higher frequency data should help.
But taxonomy shows does not reduce impacts of breaks:
Castle and Hendry (2007).

(4) Appropriate model form entails non-linear reactions.

Test for general form of non-linearity.

Castle and Hendry (2005b) use low-dimensional, orthogonalized-representation of polynomial functions. Power against up to quintics and inverses thereof.

(5) **Non-linear model selection—many sub-problems:**

(A) specify general form of non-linearity;

(B) collinearity between non-linear functions;

(C) non-normality: non-linear functions capture outliers;

(D) excess numbers of irrelevant variables;

(E) potentially more variables than observations.

Resolving model selection

Have solutions to all five sub-problems:

- (A) up to cubic polynomials, followed by encompassing tests against specific 'ogive' forms;
- (B) double demeaning removes key collinearity;
- (C) remove outliers by impulse saturation;
- (D) super-conservative Gets strategy;
- (E) multi-stage 'combinatorial selection'.

Automatic algorithm proposed in

Castle and Hendry (2005a)

Insurance policies

(6) Robust forecasting devices.

‘Insurance’ after a break to mitigate systematic failure.

**Clements and Hendry (1999); Hendry (2006):
new explanation for success of naive devices.**

(7) Improved intercept corrections.

‘Set on track’ at the forecast origin, while smoothing recent corrections. **Hendry and Santos (2005), Hendry and Reade (2006).**

(8) Pooling of forecasts.

‘Model averaging’ can go seriously wrong, but improved by Gets model selection. **Hendry and Reade (2004).**

Not going to discuss ‘insurance’ aspect in (6)–(8) today.

Route map

(A) Introduction.

(B) **Predictability and forecastability.**

(C) Taxonomies of forecast errors.

(D) Impulse saturation tests.

(E) Formulating and modelling non-linearity.

Conclusion.

Unpredictability

Non-degenerate ν_t is unpredictable wrt \mathcal{I}_{t-1} over \mathcal{T} if:

$$D_{\nu_t}(\nu_t | \mathcal{I}_{t-1}) = D_{\nu_t}(\nu_t) \quad \forall t \in \mathcal{T}. \quad (1)$$

Property of ν_t in relation to \mathcal{I}_{t-1} intrinsic to ν_t ;
 \mathcal{T} may be a singleton ($\{t\}$).

Necessary that ν_t is unpredictable in mean and variance:

$$E_t[\nu_t | \mathcal{I}_{t-1}] = E_t[\nu_t] \quad \text{and} \quad V_t[\nu_t | \mathcal{I}_{t-1}] = V_t[\nu_t].$$

Former does not imply latter, or *vice versa*.

Take $E_t[\nu_t] = \mathbf{0} \quad \forall t$.

Predictability requires combinations with \mathcal{I}_{t-1} as in:

$$y_t = \phi_t(\mathcal{I}_{t-1}, \nu_t) \quad (2)$$

Properties of unpredictability

\mathbf{y}_t depends on information set and innovation, so:

$$D_{\mathbf{y}_t}(\mathbf{y}_t \mid \mathcal{I}_{t-1}) \neq D_{\mathbf{y}_t}(\mathbf{y}_t) \quad \forall t \in \mathcal{T}. \quad (3)$$

Predictability not invariant to inter-temporal transforms.
Two most relevant special cases of (2) are:

$$\mathbf{y}_t = \mathbf{f}_t(\mathcal{I}_{t-1}) + \boldsymbol{\nu}_t \quad (4)$$

$$\mathbf{y}_t = \boldsymbol{\nu}_t \odot \boldsymbol{\varphi}_t(\mathcal{I}_{t-1}) \quad (5)$$

\odot is element by element multiplication, $y_{i,t} = \nu_{i,t} \varphi_{i,t}(\mathcal{I}_{t-1})$.

\mathbf{y}_t in (4) is predictable even if $\boldsymbol{\nu}_t$ is not:

$$E_t[\mathbf{y}_t \mid \mathcal{I}_{t-1}] = \mathbf{f}_t(\mathcal{I}_{t-1}) \neq E_t[\mathbf{y}_t].$$

‘Events’ which help predict \mathbf{y}_t must have happened.

Prediction from reduced information

Use $\mathcal{J}_{t-1} \subset \mathcal{I}_{t-1}$ to predict $\mathbf{y}_t = \boldsymbol{\nu}_t + \mathbf{f}_t(\mathcal{I}_{t-1})$:
less accurate, but unbiased predictions result.

Since $\mathbf{E}_t[\boldsymbol{\nu}_t | \mathcal{I}_{t-1}] = \mathbf{0}$:

$$\mathbf{E}_t[\boldsymbol{\nu}_t | \mathcal{J}_{t-1}] = \mathbf{0},$$

then:

$$\mathbf{E}_t[\mathbf{y}_t | \mathcal{J}_{t-1}] = \mathbf{E}_t[\mathbf{f}_t(\mathcal{I}_{t-1}) | \mathcal{J}_{t-1}] = \mathbf{g}_t(\mathcal{J}_{t-1}).$$

Let $\mathbf{e}_t = \mathbf{y}_t - \mathbf{g}_t(\mathcal{J}_{t-1})$, then $\mathbf{E}_t[\mathbf{e}_t | \mathcal{J}_{t-1}] = \mathbf{0}$, so \mathbf{e}_t is a mean innovation wrt \mathcal{J}_{t-1} , but not wrt \mathcal{I}_{t-1} :

$$\mathbf{V}_t[\mathbf{e}_t] > \mathbf{V}_t[\boldsymbol{\nu}_t].$$

Helps sustain ‘factor forecasting’, and ‘pooling of forecasts’.
Counter to claim that: ‘simple models do best’.

Simplicity is confounded with robustness.

Predicting breaks

Two important information sets for change in y_{T+1} :

(A) \mathcal{I}_T which **enters** $f_{T+1}(\mathcal{I}_T)$;

(B) \mathcal{K}_T which **explains shifts**: $f_{T+1}(\cdot) \neq f_T(\cdot)$.

Former are standard economic forces: (e.g.) money demand depends on incomes, prices, interest rates.

Latter are factors that shift relationships:

legislation, financial innovation, political factors.

Shift altered money demand at same incomes, prices, interest rates.

Predicting breaks requires: existence of \mathcal{K}_T ;

knowledge of contents of \mathcal{K}_T ; observing \mathcal{K}_T at time T ;

knowledge of how \mathcal{K}_T shifts $f_{T+1}(\cdot)$.

Practical difficulties are immense.

Predictability and forecastability

Learning-adjusted interest rate on retail sight deposits at banks following Banking Act of 1984.

$$R_{o,t} = w_t \cdot R_{r,t}$$

w_t is weighting function representing agents' learning about interest-bearing retail sight deposits:

$$w_t = (1 + \exp[\alpha - \beta(t - t^* + 1)])^{-1} \quad (6)$$

for $t \geq t^*$, zero otherwise, when $t^* = 1984(3)$.

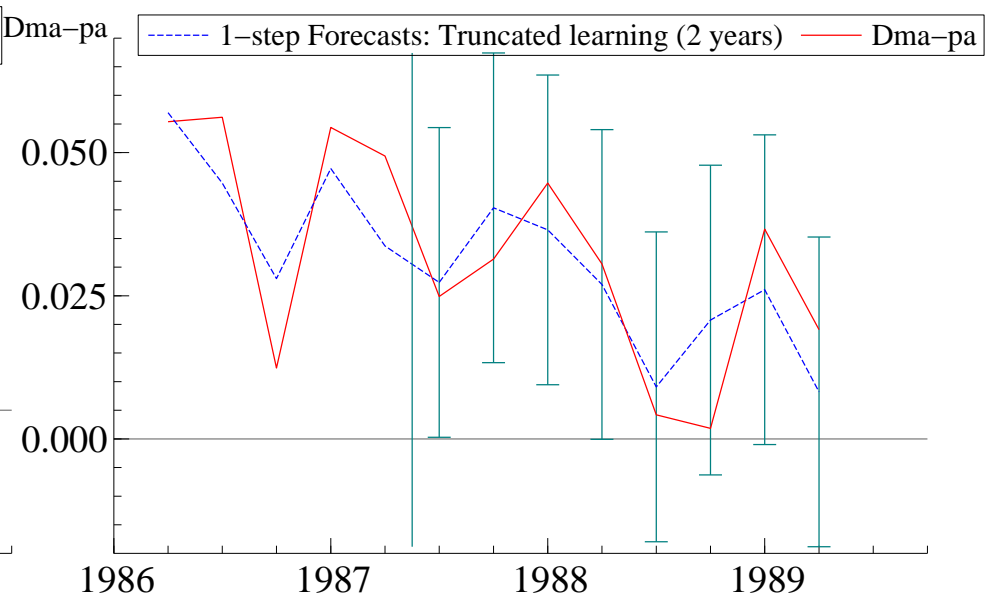
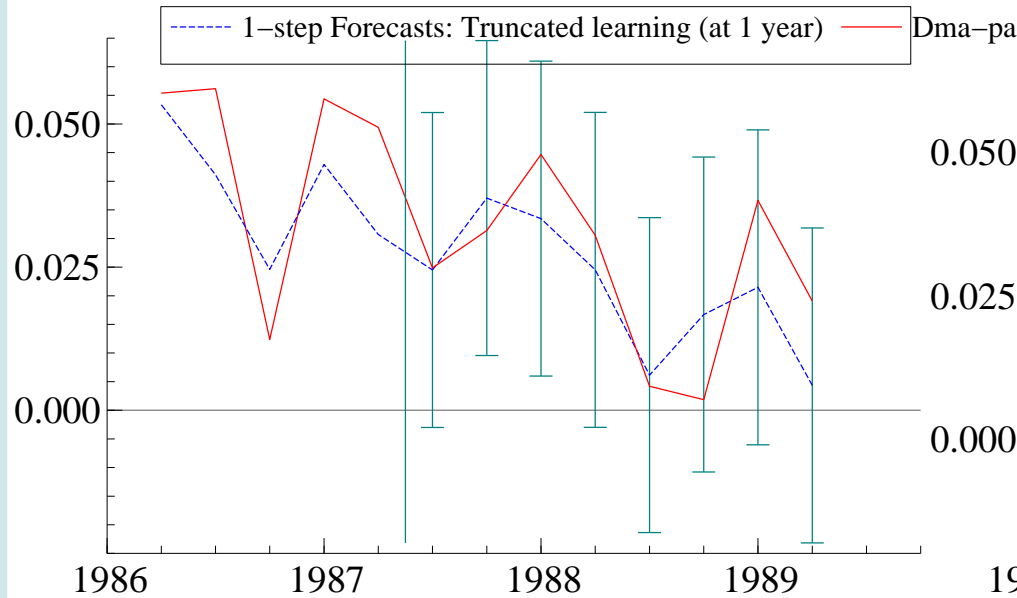
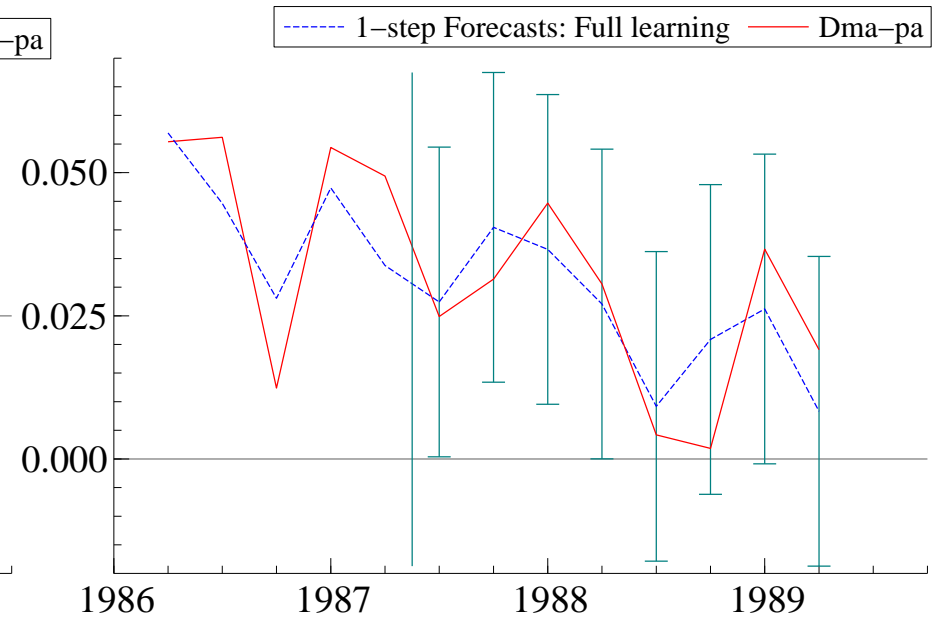
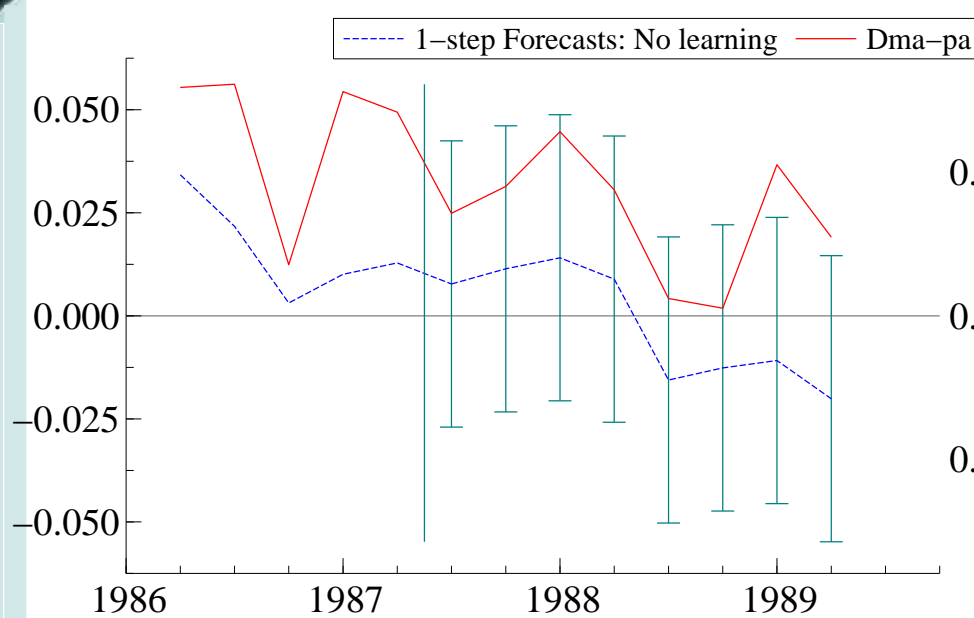
α, β estimated as in Hendry and Ericsson (1991).

Figure 13 shows four stages of (6):

(a) none; (b) full (5 years); (c) after 1 year's information;
(d) after 2 year's information.

Find (c) is enough to forecast rest of impact

M1 forecasts at 4 stages of 'learning'



Route map

(A) Introduction.

(B) Predictability and forecastability.

(C) **Taxonomies of forecast errors.**

(D) Impulse saturation tests.

(E) Formulating and modelling non-linearity.

Conclusion.

Forecast-error taxonomies

Taxonomies decompose forecast error into components: each zero if that problem is absent—all could be non-zero.

(ia)[†] equilibrium-mean shift

(ib) slope shift

(iia)^{*,†?} equilibrium-mean mis-specification

(iib) slope mis-specification

(iii)[†] forecast-origin uncertainty

(iv)^{*} estimation uncertainty

(v)^{*} error accumulation

Most components not estimable from sample evidence; but three are (*).

Most mean zero, but three need not be (†).

Reveals major sources of biases and variances.

Comparative taxonomies

Compare different methods for same setting:
disaggregation over variables (aggregate v. components);
disaggregation over time (quarterly v. monthly).

Taxonomy reveals only benefit from estimation uncertainty versus mis-specification trade-off, even if DGP is disaggregate: no advantage for breaks or innovation error if forecast aggregate.

y_t is vector of n disaggregates. In-sample DGP is $I(0)$ VAR:

$$y_t = \mu + \Gamma y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim \text{ID}[0, \Omega] \text{ for } t = 1, \dots, T \quad (7)$$

Break at the forecast origin T (but stays $I(0)$):

$$y_{T+h} = \mu^* + \Gamma^* y_{T+h-1} + \epsilon_{T+h} \text{ for } h = 1, \dots, H \quad (8)$$

Disaggregate DGP

Taking expectations in (7) under stationarity:

$$E[y_t] = \mu + \Gamma E[y_{t-1}] = \mu + \Gamma \phi_y = \phi_y$$

so $\phi_y = (\mathbf{I}_n - \Gamma)^{-1} (\mu)$ and hence:

$$y_t - \phi_y = \Gamma (y_{t-1} - \phi_y) + \epsilon_t. \quad (9)$$

$$y_{T+1} - \phi_y^* = \Gamma^* (y_T - \phi_y^*) + \epsilon_{T+1}. \quad (10)$$

From (9), forecast vector is:

$$\hat{y}_{T+1|T} = \hat{\phi}_y + \hat{\Gamma} (y_T - \hat{\phi}_y) \quad (11)$$

$y_t = \omega'_t y_t$ is aggregate, with weights ω_t .

From $T + 1$ on, forecast errors $\hat{\epsilon}_{T+1|T} = y_{T+1} - \hat{y}_{T+1|T}$ are:

$$\omega' \hat{\epsilon}_{T+1|T} = \omega' \phi_y^* - \omega' \hat{\phi}_y + \omega' \Gamma^* (y_T - \phi_y^*) - \omega' \hat{\Gamma} (y_T - \hat{\phi}_y) + \epsilon_{T+1}. \quad (12)$$

Aggregated disaggregate taxonomy

Let $\text{plim}_{T \rightarrow \infty} \hat{\Gamma} = \Gamma_p$ and $\text{plim}_{T \rightarrow \infty} \hat{\phi}_y = \phi_{y,p}$ with
 $\hat{\Gamma} = \Gamma_p + \Delta_{\Gamma}$, $\hat{\phi}_y = \phi_{y,p} + \delta_y$ etc.

Aggregated disaggregate forecast-error taxonomy

$$\begin{aligned}
 & \omega' \hat{\epsilon}_{T+1|T} = \\
 & \omega' (\mathbf{I}_n - \Gamma^*) (\phi_y^* - \phi_y) && (ia) \text{ mean change} \\
 & + \omega' (\Gamma^* - \Gamma) (\mathbf{y}_T - \phi_y) && (ib) \text{ slope change} \\
 & + \omega' (\mathbf{I}_n - \Gamma_p) (\phi_y - \phi_{y,p}) && (iia) \text{ mean mis-specification} \\
 & + \omega' (\Gamma - \Gamma_p) (\mathbf{y}_T - \phi_y) && (iib) \text{ slope mis-specification} \\
 & - \omega' (\mathbf{I}_n - \Gamma_p) \delta_y && (iiia) \text{ equilibrium-mean estimation} \\
 & - \omega' \left[\mathbf{I}_n \otimes (\mathbf{y}_T - \phi_{y,p})' \right] \Delta_{\Gamma}^{\nu} && (iiib) \text{ slope estimation} \\
 & + \omega' \Delta_{\Gamma} \delta_y && (iv) \text{ covariance interaction} \\
 & + \omega' \epsilon_{T+1} && (v) \text{ innovation error.}
 \end{aligned}$$

(13)

Derived aggregate taxonomy

Derived aggregate forecast-error taxonomy

$$\begin{aligned} \tilde{\nu}_{T+1|T} &\simeq \\ &\omega' (\phi_y^* - \phi_y) - \omega' \Gamma^* (\phi_y^* - \phi_y) && (Ia) \text{ mean change} \\ &+ \omega' (\Gamma^* - \Gamma) (\mathbf{y}_T - \phi_y) && (Ib) \text{ slope change} \\ &+ (\tau - \tau_p) && (IIa) \text{ mean mis-specification} \\ &+ \omega' (\Gamma - \kappa_p \mathbf{I}_n) (\mathbf{y}_T - \phi_y) && (IIb) \text{ slope mis-specification} \\ &+ (\tau_p - \tilde{\tau}) && (IIIa) \text{ equilibrium-mean estimation} \\ &+ (\kappa_p - \tilde{\kappa}) \omega' (\mathbf{y}_T - \phi_y) && (IIIb) \text{ slope estimation} \\ &+ \omega' \epsilon_{T+1} && (IV) \text{ innovation error.} \end{aligned} \quad (14)$$

Four conclusions from (14) and (13).

(Ia)+(Ib) identical to (ia)+(ib):

forecast-origin shifts not changed by aggregation.

Innovation errors in (IV) and (iv) also identical.

Equilibrium-mean mis-specification unlikely in both taxonomies if in-sample DGP constant.

Route map

- (A) Introduction.
- (B) Predictability and forecastability.
- (C) Taxonomies of forecast errors.
- (D) **Impulse saturation tests.**
- (E) Formulating and modelling non-linearity.
- Conclusion.**

Impulse saturation tests

Normality important for model selection:
null rejection frequency above nominal when non-linear terms capture extreme observations.

Solution of indicator saturation proposed:
also removes earlier breaks.

Still researching properties under alternative;
rules for partitioning and algorithms to implement;
characterizing the selected estimated impulses.

Testing non-constancy by adding complete set of impulse indicators $\{1_{\{t\}}, t = 1, \dots, N\}$.

Using general-to-specific, know null distribution of mean in location-scale IID model $x_j \sim \text{IID}[\mu, \sigma_\epsilon^2]$ after adding $\{1_{\{t\}}\}$.

Impulse saturation theorem

If x_1, \dots, x_N are symmetric IID $[\mu, \sigma_\epsilon^2]$, then the estimator $\tilde{\mu}$:

$$\tilde{\mu} = \frac{\sum_{i=1}^{0.5N} x_i 1_{\{|x_i - \bar{x}_2| \leq c_\alpha\}} + \sum_{j=0.5N+1}^N x_j 1_{\{|x_j - \bar{x}_1| \leq c_\alpha\}}}{\sum_{i=1}^{0.5N} 1_{\{|x_i - \bar{x}_2| \leq c_\alpha\}} + \sum_{j=0.5N+1}^N 1_{\{|x_j - \bar{x}_1| \leq c_\alpha\}}}, \quad (15)$$

is distributed as:

$$N^{1/2} (\tilde{\mu} - \mu) \rightarrow N [0, \sigma_\epsilon^2 \sigma_\mu^2], \quad (16)$$

where:

$$\sigma_\mu^2 = \frac{1}{P(c_\alpha)} \left(1 + 4c_\alpha f(c_\alpha) - \frac{2c_\alpha f(c_\alpha)}{P(c_\alpha)} - \frac{4c_\alpha^2 f(c_\alpha)^2}{P(c_\alpha)} \right) \quad (17)$$

when:

$$P(c_\alpha) = P(|\hat{u}| \leq c_\alpha \sigma_\epsilon) = \int_{-c_\alpha}^{c_\alpha} f(u) du, \quad (18)$$

measures the impact of truncating the residuals.

Several combinations

Three-steps: half indicators added, all significant recorded; then other half examined; finally, two retained sets of indicators combined. Then:

$$N^{1/2} (\tilde{\mu} - \mu) \rightarrow N [0, \sigma_{\epsilon}^2 \sigma_{\mu}^2], \quad (19)$$

Average null retention rate is αN , at significance level α : for $\alpha = 0.01$, then $0.01N$ indicators will be retained.

Alternative splits, such as $N/3$, $N/4$, do not affect null retention rate. **Feasible algorithm exists: see Hendry, Johansen and Santos (2004).**

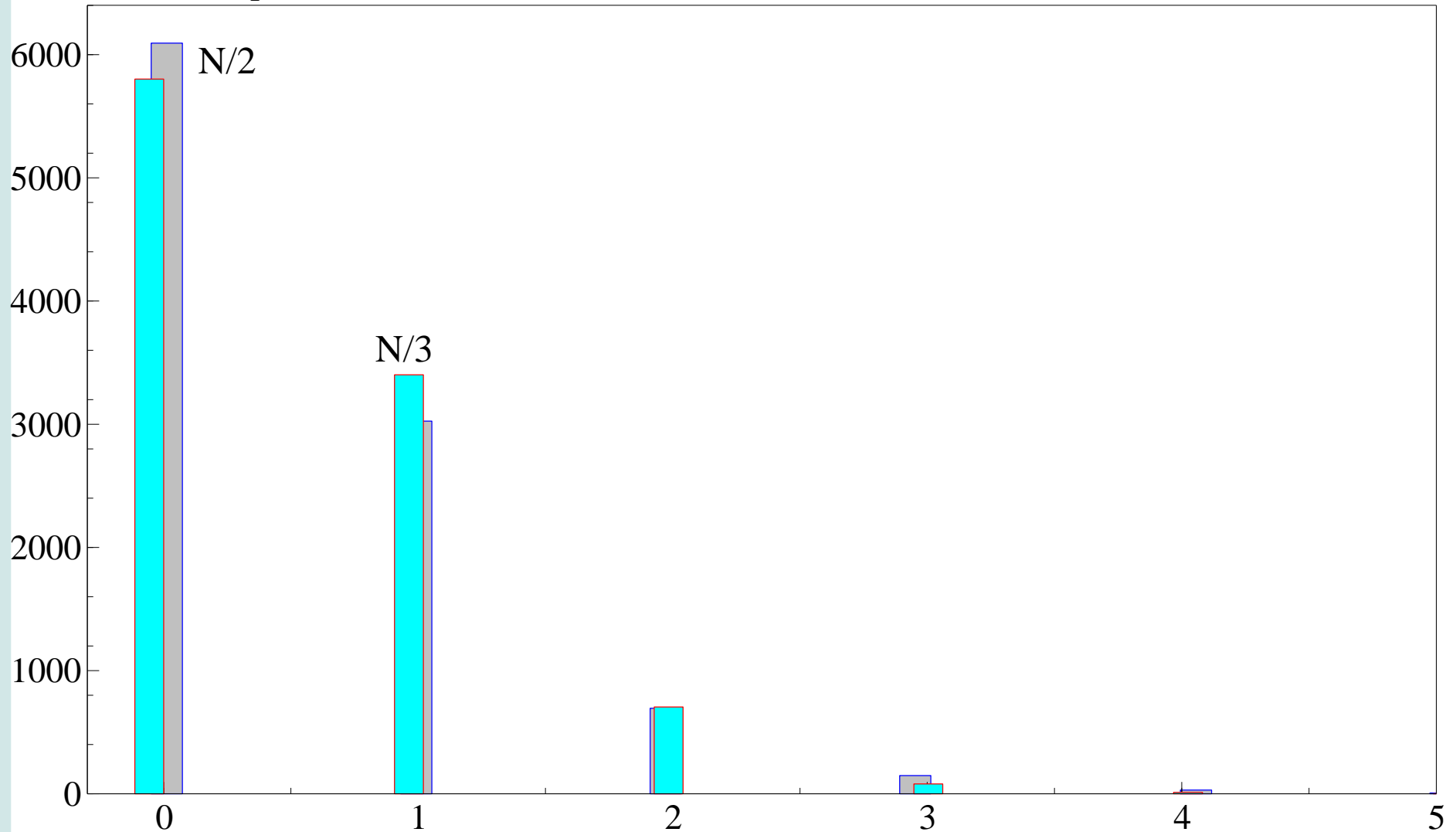
Importantly: can investigate several combinations:

$N/2$; that randomly; $N/3$ split and random etc.

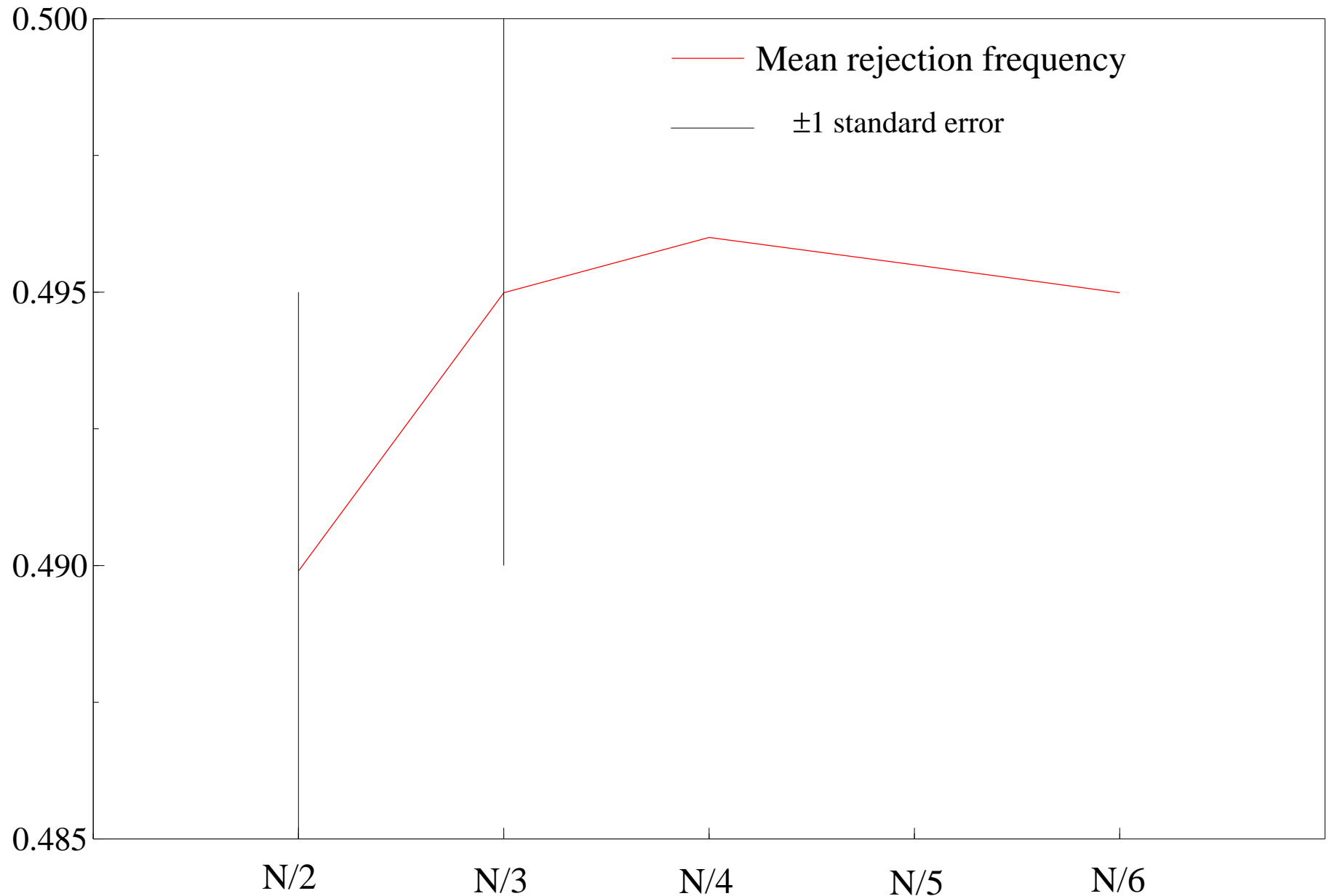
No impact on null rejection frequency: even if do all of them and combine.

Null retention for $\alpha = 1\%$

Numbers of impulses retained: $N/2$ v $N/3$ at $N=50$



Null rejection for $\alpha = 1\%$



Summary

Hendry et al. (2004) have shown:

$\tilde{\mu}$ is unbiased;

σ_{μ} is accurate in finite samples:

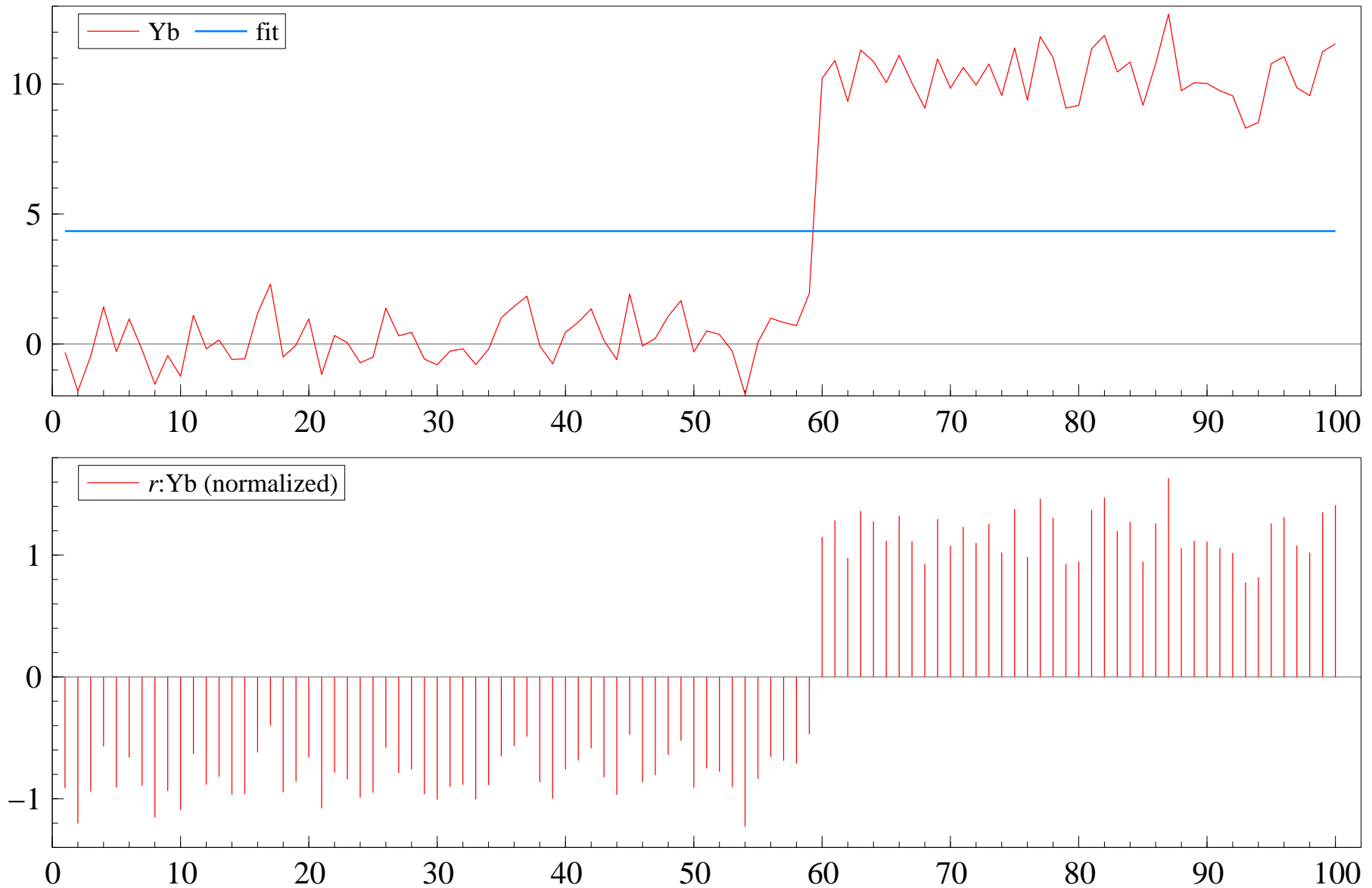
can approximately estimate by $f(\cdot) = \phi(\cdot)$ and $P(c_{\alpha}) \simeq 1 - \alpha$,

and can bias correct $\tilde{\sigma}^2$ if desired;

$\alpha \times N$ impulses retained by chance.

Important difference between outlier detection and impulse saturation: see figure 27.

Absence of outliers despite a break



Route map

- (A) Introduction.
- (B) Predictability and forecastability.
- (C) Taxonomies of forecast errors.
- (D) Impulse saturation tests.
- (E) **Formulating and modelling non-linearity** .
- Conclusion.**

Testing for non-linearity

Consider a general relation:

$$y_t = f(x_{1,t}, \dots, x_{n,t}) + \epsilon_t$$

where the linear approximation is:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_n x_{n,t} + e_t = \beta_0 + \boldsymbol{\beta}' \mathbf{x}_t + e_t$$

Mean-value theorem suggests testing for the significance of the quadratic term when added to the linear approximation.

$$y_t = \beta_0 + \boldsymbol{\beta}' \mathbf{x}_t + \boldsymbol{\delta}'_1 \mathbf{w}_t + e_t, \quad \text{where } \mathbf{w}_t = (\mathbf{x}_t \mathbf{x}'_t)^{v_e}$$

Close to White (1980)–test for heteroskedasticity.

Drawbacks

The main drawbacks of the test are:

Drawbacks

The main drawbacks of the test are:

- **High dimensionality**

Drawbacks

The main drawbacks of the test are:

- High dimensionality
- Potential high collinearity between elements of w_t

Drawbacks

The main drawbacks of the test are:

- High dimensionality
- Potential high collinearity between elements of w_t
- Departure from linearity may be better reflected in third derivative

Drawbacks

The main drawbacks of the test are:

- High dimensionality
- Potential high collinearity between elements of w_t
- Departure from linearity may be better reflected in third derivative

Alternatively use eigenvectors of variance matrix.

$$\mathbf{x}_t \sim D_n [\boldsymbol{\mu}, \boldsymbol{\Omega}],$$

$$\boldsymbol{\Omega} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}'; \mathbf{H}'\mathbf{H} = \mathbf{I}_n \quad \mathbf{z}_t^* = \mathbf{H}'\mathbf{x}_t; \Rightarrow \mathbf{z}_t^* \sim D_n [\mathbf{H}'\boldsymbol{\mu}, \boldsymbol{\Lambda}];$$

$$z_{i,t} = \frac{z_{i,t}^* - \overline{z_{i,t}^*}}{\sqrt{\lambda_i}} \Rightarrow \mathbf{z}_{t, \text{app}} \sim D_n [\mathbf{0}, \mathbf{I}].$$

Improved testing

For $u_{i,t} = z_{i,t}^2$, under the null, for fixed regressors and $e_t \sim \text{IN} [0, \sigma_e^2]$, the test of $\delta_1 = 0$ in:

$$y_t = \beta_0 + \beta' \mathbf{x}_t + \delta_1' \mathbf{u}_t + e_t$$

is an exact **F**-test with **n** degrees of freedom.

Compares to $\frac{n(n+1)}{2}$ degrees of freedom for the original test.

Including $\sum_{i=1}^n \delta_{2,i} z_{i,t}^3$ results in an **F**-test of $\delta_1 = \delta_2 = 0$ with **2n** degrees of freedom.

Solved:

- ✓ High dimensionality
- ✓ Potential high collinearity between elements of \mathbf{w}_t
- ✓ Departures from linearity in third derivative

Non-linearity testing

If functional form and set of relevant variables unknown, Index test has power to reject false null in a range of circumstances

Outperforms White's test for large n

Selection prior to implementing test hazardous if linear term irrelevant but enters DGP non-linearly

Collinearity can be beneficial as products of terms proxy polynomials

Power against inverse polynomials as high collinearity between variables and their inverses

Easily implemented and designed for large, potentially collinear GUMs

Formulating non-linearity

1] Commence with general polynomial approximation:

$$y_t = \beta' \mathbf{X}_t + \alpha' g(\mathbf{X}_t) + v_t. \quad \text{for } t = 1, \dots, T.$$

$\mathbf{X}_t = (n \times 1)$ vector of potentially relevant variables

$g(\mathbf{X}_t) = (m \times 1)$ vector of non-linear transformations.

2] Undertake selection (using techniques for more variables than observations)

3] Test model against preferred non-linear functional form

Problems solved:

- ✓ Reparameterize to mimic orthogonal representation
- ✓ Remove extreme observations using 'indicator saturation' techniques
- ✓ Develop a 'super-conservative' selection strategy

Non-linear Functions: Polynomials

What class of functions captures non-linearity inherent in economic data?

A number explored: **we chose polynomials**

Polynomials approximate STR models (which may nest many regime-switching models, neural networks, etc.):

$$y_t = \beta' \mathbf{X}_t + (\theta' \mathbf{X}_t) G(s_t; \gamma, c) + u_t, \quad u_t \sim \text{IN} [0, \sigma_u^2]$$

Logistic transition function:

$$G(s_t; \gamma, c) = \left[1 + \exp \left\{ -\gamma \left(\frac{s_t - c}{\hat{\sigma}_s} \right) \right\} \right]^{-1}$$

s_t = transition variable

c = switching threshold parameter

γ = steepness parameter

Polynomials as approximations

To approximate LSTR(1) use 3rd order Taylor expansion:

$$y_t \simeq \beta' \mathbf{X}_t + (\theta' \mathbf{X}_t) \left[\frac{1}{2} + \frac{z_t}{4} - \frac{z_t^3}{48} \right] + v_t$$

where $z_t = \gamma \left(\frac{s_t - c}{\hat{\sigma}_s} \right)$, which can be estimated as:

$$y_t \simeq \alpha'_1 \mathbf{X}_t + \alpha'_2 \mathbf{X}_t s_t + \alpha'_3 \mathbf{X}_t s_t^2 + \alpha'_4 \mathbf{X}_t s_t^3 + v_t$$

For \mathbf{X}_t scalar:

$$\alpha_1 = \beta + \frac{\theta}{2} - \frac{\theta \gamma c}{4 \hat{\sigma}_s} + \frac{\theta \gamma^3 c^3}{48 \hat{\sigma}_s^3}$$

$$\alpha_2 = \frac{\theta \gamma}{4 \hat{\sigma}_s} - \frac{3 \theta \gamma^3 c^2}{48 \hat{\sigma}_s^3}$$

$$\alpha_3 = \frac{3 \theta \gamma^3 c}{48 \hat{\sigma}_s^3} \quad \alpha_4 = -\frac{\theta \gamma^3}{48 \hat{\sigma}_s^3}$$

General-to-specific modelling

In practice we don't know relevant variables, transition variable or lag lengths.

Therefore start from GUM and do Gets.

$$y_t = \sum_{i=1}^N \delta_i W_{i,t} + \sum_{i=1}^N \sum_{j=1}^M \kappa_{ij} W_{i,t} z_{j,t} + \sum_{i=1}^N \sum_{j=1}^M \lambda_{ij} W_{i,t} z_{j,t}^2 + \sum_{i=1}^N \sum_{j=1}^M \psi_{ij} W_{i,t} z_{j,t}^3$$

N potential regressors, \mathbf{W}_t

M potential transition variables, \mathbf{z}_t

Problems

- Correlations between functions
- Estimation issues – regime switches

Functional form approximation

k variables retained ($k \leq n$) and one transition variable, s_1 .
Test the postulated functional form by:

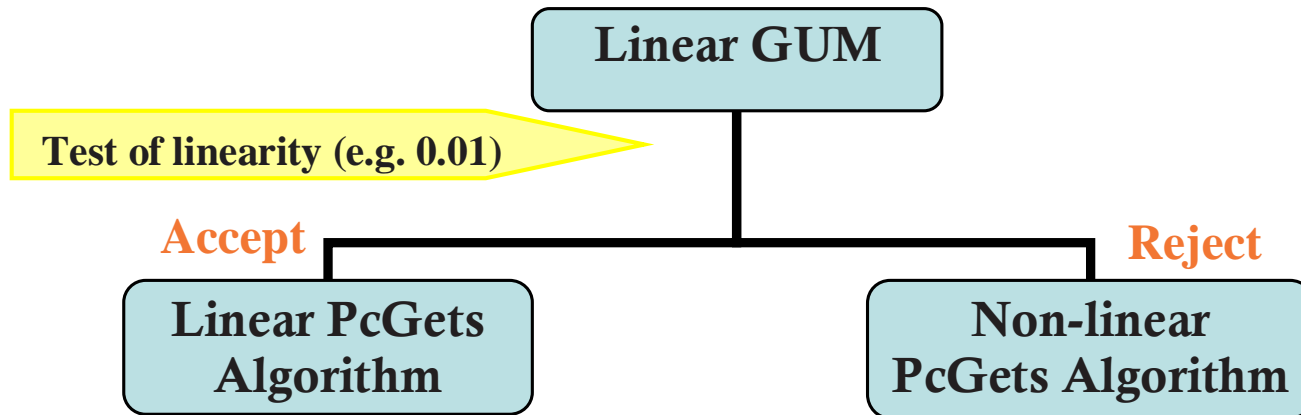
$$H_0 : \kappa = \mu = \psi = 0,$$

for:

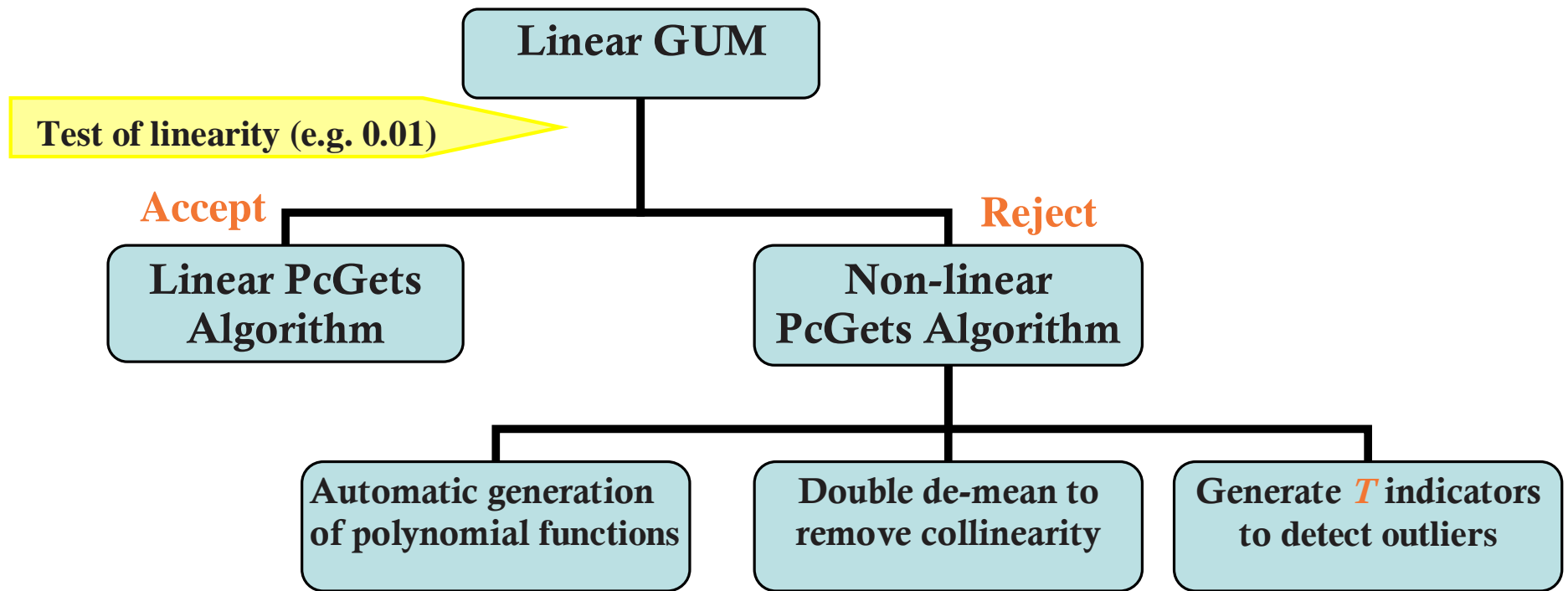
$$y_t = \sum_{i=1}^k \tau_i W_{i,t} + \sum_{i=1}^k \kappa_i W_{i,t} s_{1,t} + \sum_{i=1}^k \mu_i W_{i,t} s_{1,t}^2 + \sum_{i=1}^k \psi_i W_{i,t} s_{1,t}^3 \\ + \sum_{i=1}^k \left(\tilde{\theta}_i W_{i,t} \right) \left[1 + \exp \left\{ -\tilde{\gamma} \left(\frac{s_{1,t} - \tilde{c}}{\hat{\sigma}_{s_1}} \right) \right\} \right]^{-1} + \eta_t$$

Solves identification problems of **Granger and Teräsvirta (1993)**, while concluding with LSTAR model if best representation.

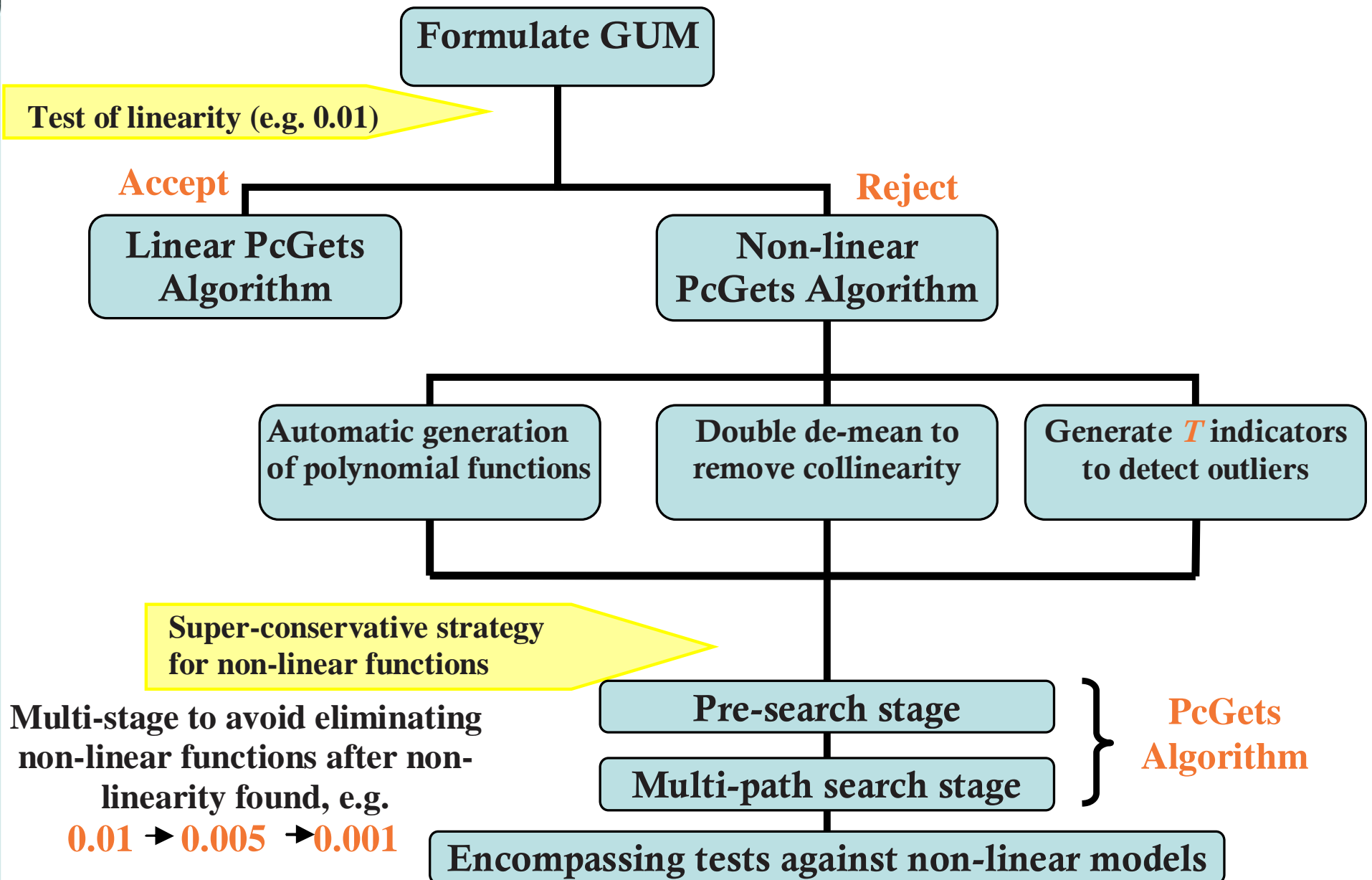
Non-linear Algorithm



Non-linear Algorithm



Non-linear Algorithm



Conclusions

Whether breaks are predictable from relevant information available at the forecast origin remains unknown as yet. **But progress in developing forecasting models; and methods of testing for and selecting such models.**

Predictability theory: 2 information sets, regular and shifts; model latter as non-linear ogive.

Forecast-error taxonomies: what factors matter.
Surprise that breaks affect (dis)aggregates the same.

Handle non-normality by impulse saturation: size is controlled, distributions well behaved, and power for a range of interesting issues.

Applies to more variables than observations and hence invaluable in non-linear modelling.

Conclusions II

Extend automatic Gets algorithm for non-linear functions:
focus on polynomials

- Test of functional form
- Operational rules to orthogonalize
- Indicator saturation to remove extreme observations
- Number of potential non-linear variables is large:
Super-conservative strategy
- Approximate a wide range of non-linear models

Selection of non-linear models analogous to linear models
Approximations capture unknown non-linear functional form

Non-linear capability feasible for Gets

References

Bibliography

- Castle, J. L., and Hendry, D. F. (2005a). Extending the boundaries of PcGets: Non-linear models. Mimeo, Economics Department, Oxford University.
- Castle, J. L., and Hendry, D. F. (2005b). A low-dimension collinearity-robust test for non-linearity. Mimeo, Economics Department, Oxford University.
- Castle, J. L., and Hendry, D. F. (2007). Forecasting UK inflation: Empirical evidence on robust forecasting devices. In Rapach, D. E., and Wohar, M. E. (eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty*. Amsterdam: Elsevier. Forthcoming.
- Clements, M. P., and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.
- Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.
- Clements, M. P., and Hendry, D. F. (2005). Guest editors' introduction: Information in economic forecasting. *Oxford Bulletin of Economics and Statistics*, **67**, 713–753.
- Clements, M. P., and Hendry, D. F. (2006). Forecasting with breaks in data processes. In Elliott, G., Granger, C. W. J., and Timmermann, A. (eds.), *Handbook of Econometrics on Forecasting*, pp. 605–657. Amsterdam: Elsevier.
- Granger, C. W. J., and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.
- Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics*, **135**, 399–426. Special Issue in Honor of Clive Granger.
- Hendry, D. F., and Ericsson, N. R. (1991). Modeling the demand for narrow money in the United Kingdom and the United States. *European Economic Review*, **35**, 833–886.
- Hendry, D. F., Johansen, S., and Santos, C. (2004). Selecting a regression saturated by indicators. Unpublished paper, Economics Department, University of Oxford.
- Hendry, D. F., and Reade, J. (2004). Problems with model averaging, and a better alternative. Working paper, Economics Department, Oxford University.

Rerouting

- (A) **Introduction.**
 - (B) **Predictability and forecastability.**
 - (C) **Taxonomies of forecast errors.**
 - (D) **Impulse saturation tests.**
 - (E) **Formulating and modelling non-linearity.**
- Conclusion.**

Undetectable breaks

DGP is VAR(1):

$$\mathbf{y}_t = \boldsymbol{\phi} + \boldsymbol{\Pi} \mathbf{y}_{t-1} + \mathbf{u}_t, \quad t = 1, \dots, T_1$$

$$\mathbf{y}_t = \boldsymbol{\phi}^* + \boldsymbol{\Pi}^* \mathbf{y}_{t-1} + \mathbf{u}_t, \quad t = T_1 + 1, \dots, T$$

where: $\mathbf{u}_t \sim \text{IN}[\mathbf{0}, \boldsymbol{\Omega}]$ when $\omega_{i,i} = 0.01^2$.

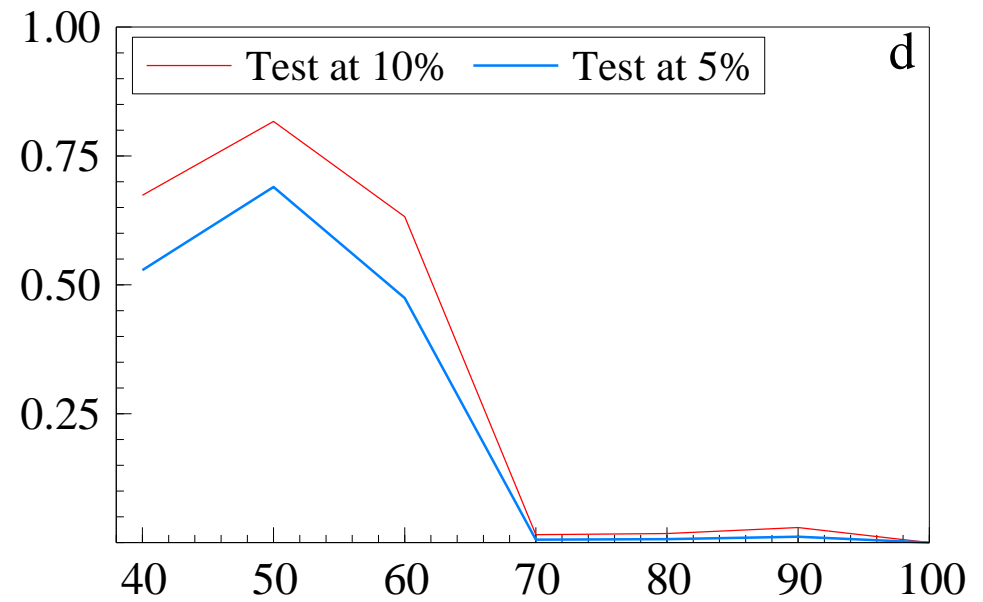
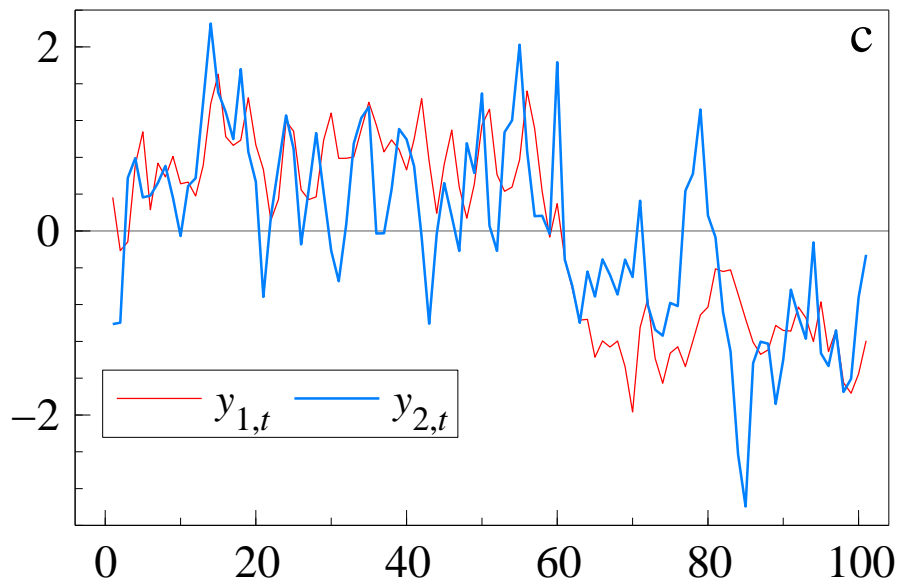
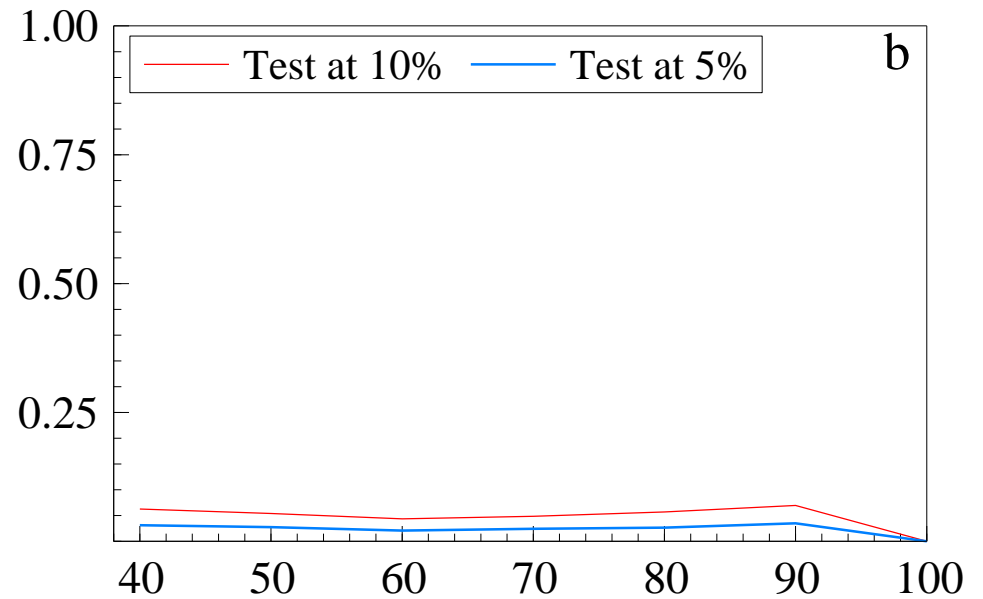
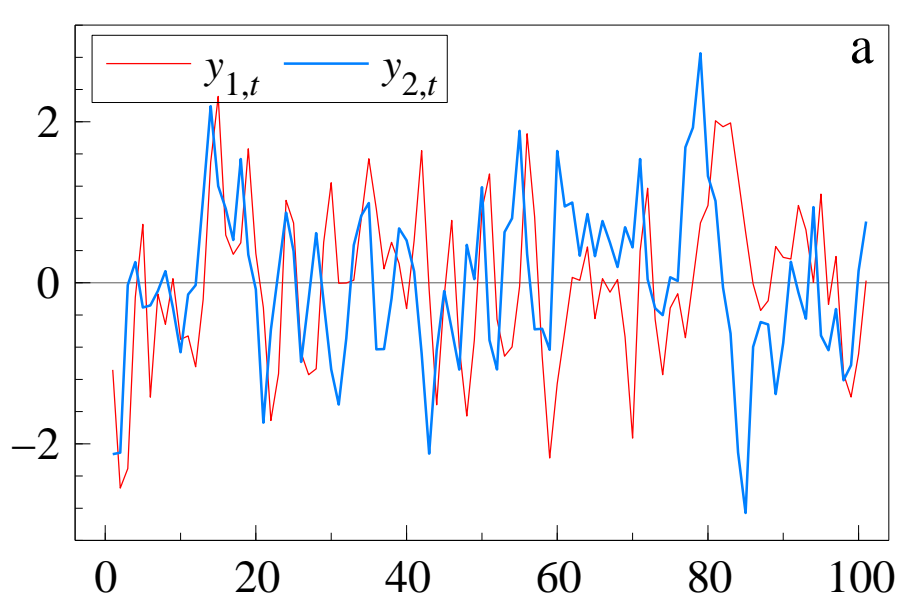
$$\boldsymbol{\Pi} = \begin{pmatrix} 0.5 & 0.20 \\ -0.20 & 0.4 \end{pmatrix}, \quad \boldsymbol{\phi} = \begin{pmatrix} 1.0 \\ -1.0 \end{pmatrix}$$

Then at $T_1 = 0.6T$, all $\pi_{i,j}$ change by 30ω , and ϕ_i by 97ω :

$$\boldsymbol{\Pi}^* = \begin{pmatrix} 0.2 & 0.50 \\ -0.50 & 0.7 \end{pmatrix}, \quad \boldsymbol{\phi}^* = \begin{pmatrix} 1.97 \\ -.03 \end{pmatrix}$$

Fundamentally changed process: fig. 46a shows data; fig. 46b, Chow rejections.

Chow test rejection frequencies



Detectable breaks

But with same Π , small change in ϕ by just 3ω to:

$$\begin{pmatrix} 0.97 \\ -0.97 \end{pmatrix}$$

is easily detected, as figs. 46c,d show.

Key is long-run mean, $E[y_t] = (\mathbf{I} - \Pi)^{-1} \phi$, stays at:

$$\begin{pmatrix} 1.176 \\ -2.059 \end{pmatrix}$$

in first case, despite changes; but shifts from that to:

$$\begin{pmatrix} 1.141 \\ -1.997 \end{pmatrix}$$

in second. **Location shifts are key to detection.**

Monte Carlo experiments

The DGP is given by:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + (\delta_0 + \delta_1 x_{1,t} + \delta_2 x_{2,t}) [1 + \exp(-\gamma (x_{1,t} - c))]^{-1}$$

with $\mathbf{x}_t = \rho \mathbf{x}_{t-1} + \boldsymbol{\nu}_t$; $\eta_t \sim \text{IN}[0, 1]$, $\boldsymbol{\nu}_t \sim \text{IN}_4[0, \mathbf{I}]$.

Potential variables: $\mathbf{W}_t = x_1, x_2, x_3, x_4$

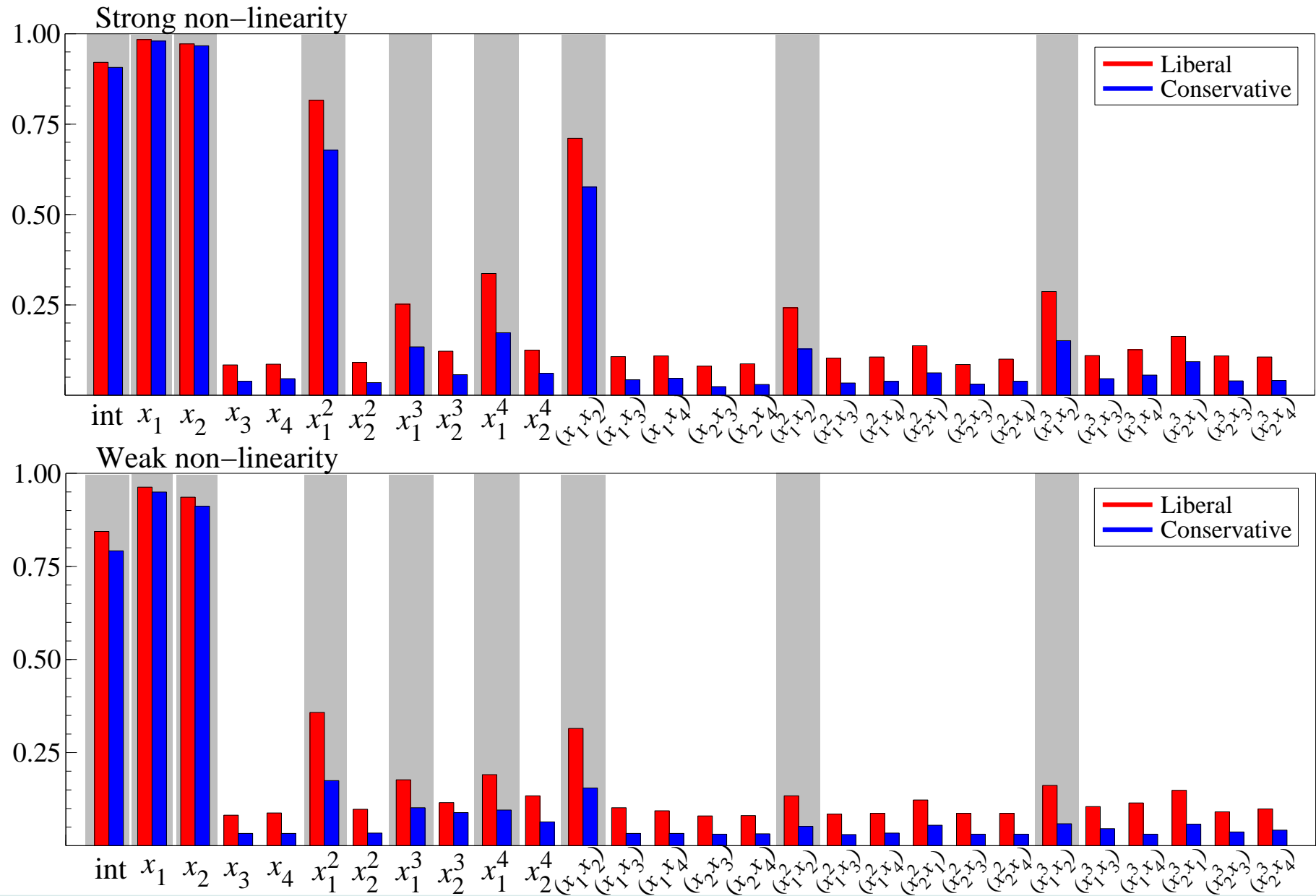
Potential transition variables: $\mathbf{z}_t = x_1, x_2$

GUM contains 28 variables: $\mathbf{1}, x_1, x_2, x_3, x_4, x_1^2, x_2^2, x_1^3, x_2^3, x_1^4, x_2^4,$
 $x_1 x_2, x_1 x_3, x_1 x_4, x_2 x_3, x_2 x_4, x_1^2 x_2, x_1^2 x_3, x_1^2 x_4, x_2^2 x_1, x_2^2 x_3, x_2^2 x_4,$
 $x_1^3 x_2, x_1^3 x_3, x_1^3 x_4, x_2^3 x_1, x_2^3 x_3, x_2^3 x_4$

Various parameter values considered:

$$\beta_0 = 0.2, \beta_1 = \beta_2 = 0.1, \delta_0 = 0.8, \delta_1 = \delta_2 = 0.8, \gamma = 4, c = 0.5, \rho = 0.8,$$
$$\beta_0 = 0.2, \beta_1 = \delta_1 = 0.3, \beta_2 = \delta_2 = 0.4, \delta_0 = 0.8, \gamma = 3, c = 0.5, \rho = 0.8$$

MC results



Size using Hermite Polynomials

