Deciding between Alternative Approaches in Macroeconomics

David F. Hendry

Institute for New Economic Thinking at the Oxford Martin School, University of Oxford

Research jointly with Jennifer Castle, Jurgen Doornik, Søren Johansen, Grayham Mizon and Bent Nielsen

ESRC/OMS Conference, 2012
[1] All macroeconomic theories are incomplete, incorrect and changeable;

[2] all macroeconomic time-series data are aggregated, inaccurate and rarely match theoretical concepts;

[3] all empirical macro-econometric models are non-constant, and

[4] mis-specified in numerous ways;

[5] economic policy often has unexpected effects different from prior analyses;

[6] macroeconomic forecasts regularly go awry:

so how to decide between alternative approaches?
Main justification of empirical macro-econometric evidence is conformity with conventionally-accepted economic theory: ‘internal credibility’ as against verisimilitude.

Partly justified by manifest inadequacy of short, dependent, and heterogeneous time-series data, often subject to extensive revision: if data are unreliable, better to trust the theory.

But theories have evolved greatly, and most previous analyses abandoned: almost self-contradictory to justify an empirical model by an invalid theory that will soon be altered.

Why is incorrect and mutable theory more reliable than data evidence?
Worries concern prevalence of non-stationarity, endogeneity, potential lack of identification, and collinearity, belief that ‘data mining’ can produce almost any desired result—
but so can theory choice by matching non-existent ‘stylized facts’, that are neither constant nor facts.

Part is a mistaken conflation of economic-theory models of human behavior with the data generation process (DGP):
huge gap between abstract theory and non-stationary evidence finessed by asserting that the model is the mechanism—
‘let’s take the model seriously’—no, let’s leave the talk.

Final part is false belief that data-based model selection is a subterfuge of scoundrels—
rather than the key to understanding the complexities of macro-economies.
Reality includes unanticipated shifts.
Reality includes unanticipated shifts
Reality includes unanticipated shifts
Both credibility & verisimilitude matter

Cannot discuss all pertinent issues here, focus on constellation of:
[1] model selection while retaining (not imposing) theory;
[2] tackling multiple location shifts; and

Forecasting addressed in Castle, Fawcett, and Hendry (2011);
and nowcasting in Castle, Fawcett, and Hendry (2009).
How to proceed?

Nest ‘theory-driven’ and ‘data-driven’ approaches: **retain theory insights unaffected by selection, but select over rival candidate variables, lags, functional forms, etc..**

Multiple breaks also rarely included in theories but can be accommodated empirically.

Multi-path search algorithm with tight critical values controls false retention of irrelevant variables at low levels yet retains all theory-based variables, irrespective of their significance.

**If theory correct and complete, distributions of parameter estimators identical to directly fitting to data: if incorrect, discover better empirical model.**

Can have more candidate variables, \( N \), than sample size \( T \).

Yet selection costless if not needed and beneficial otherwise, **precisely the opposite of current beliefs.**
Discoveries in economics mainly from theory.

But all economic theories are:
(a) incomplete; (b) incorrect; and (c) mutable.

(a) Need strong ceteris paribus assumptions: inappropriate in a non-stationary, evolving world.

(b) Consider an economic analysis which suggests:
\[ y = f(x) \]  

where the \( k \) variables \( y \) depend on \( r \) ‘explanatory’ variables \( x \) with \( m > r \) instruments \( z \).

Form of \( f(\cdot) \) in (1) depends on:
utility or loss functions of agents,
constraints they face, & information they possess.

Analyses arbitrarily assume: forms for \( f(\cdot) \), that \( f(\cdot) \) is constant, that only \( x \) matters, & that the \( z \)s are ‘exogenous’.

Yet must aggregate across heterogeneous individuals whose endowments shift over time, often abruptly.
(c) Economic analyses have not only changed our understanding, they have changed the world: from the ‘invisible hand’ in Adam Smith’s *Theory of Moral Sentiments* (1759, p.350) onwards, theory has progressed dramatically—key insights into option pricing, auctions and contracts, principal-agent and game theories, trust and moral hazard, asymmetric information, institutions: major impacts on market functioning, industrial, and even political, organization.

**But imagine imposing 1900’s economic theory in empirical research today.**

Much past applied econometrics research is forgotten: discard the economic theory that it ‘quantified’ and you discard the associated empirical evidence.

**Hence fads & fashions, ‘cycles’ and ‘schools’ in economics.**
Need to tackle all complications jointly. 
Re-frame empirical modeling as discovery process combined with theory evaluation.

Starting from $T$ observations on $N > r$ variables $w$, aim to find $\beta^*$ for $s$ lagged functions $g(w^*_t) \ldots g(w^*_{t-s})$ of a subset of $k$ variables $w^*$, jointly with $\{1_{t=t_i}\}$—indicators for breaks, outliers etc.

Embeds initial economic analysis $y = f(x)$, but in a much more general initial model.

Globally, learning must be simple to general; but locally need not be in observational disciplines.

Approach explained in Castle, Doornik, and Hendry (2011) and Hendry and Johansen (2012).
Extensions determine how well DGP is approximated

Create three extensions automatically:

(i) lag formulation to implement sequential factorization;
(ii) functional form transformations for non-linearity;
(iii) impulse-indicator saturation (IIS) for parameter non-constancy and data contamination.

(i) Create $s$ lags $w_t \ldots w_{t-s}$ to formulate general linear model:

$$y_t = \beta_0 + \sum_{i=1}^{s} \lambda_i y_{t-i} + \sum_{i=1}^{r} \sum_{j=0}^{s} \beta_{i,j} w_{i,t-j} + \epsilon_t$$

Focus on single equations, but systems can be handled.
(ii) Castle and Hendry (2011) propose cubics and exponentials of principal components $u_t$ of the $w_t$.

Let $w_t \sim D_r [\mu, \Omega]$, where $\Omega = H\Lambda H'$ with $H'H = I_r$.

Empirically: $\hat{\Omega} = T^{-1} \sum_{t=1}^{T} (w_t - \bar{w})(w_t - \bar{w})' = \hat{H}\hat{\Lambda}\hat{H}'$

so that $u_t = \hat{H}'(w_t - \bar{w})$ leading to $u_t \sim D_r [0, I]$.

Presently implemented general cubics with exponential functions.

$u^2_{i,t}; u^3_{i,t}; u_{i,t}e^{-|u_{i,t}|}$.

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

Low dimensional, with no collinearity between elements of $u_t$ yet includes most important sources of departure from linearity, e.g. asymmetry.
Non-linear functions

\[ z_i^2 \times z_i \]

\[ z_i^3 \times z_i \]

\[ z_i e^{-|z_i|} \times z_i \]
Numbers, timings and magnitudes of breaks in models usually unknown: obviously so for unknowingly omitted variables. ‘Portmanteau’ approach required to detect location shifts anywhere in sample, while also selecting over many candidate variables.

To check the null of no outliers or location shifts in a model, impulse-indicator saturation (IIS) creates complete set of indicator variables:
\[ \{1_{j=t}\} = 1 \text{ when } j = t \text{ and } 0 \text{ otherwise for } j = 1, \ldots, T: \]
need to add \( T \) impulse indicators to set of candidate variables when \( T \) observations.

Johansen and Nielsen (2009) extend IIS to both stationary and unit-root autoregressions.

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

$$T^{1/2}(\tilde{\beta} - \beta) \xrightarrow{D} N_n \left[ 0, \sigma^2 \Sigma^{-1} \Omega_\beta \right]$$

Same rate of convergence, to normal, with correct centering, despite $T$ extra indicators.

Efficiency of IIS estimator $\tilde{\beta}$ with respect to OLS $\hat{\beta}$ measured by $\Omega_\beta$ depends on $c_\alpha$ and distribution form.

Must lose efficiency under null; small loss $\alpha T$ of 1 observation at $T = 100$ if $\alpha = 1/T = 0.01$ so 99% efficient.

Potential for major gain under alternatives of breaks and/or outliers, yet can be done jointly with all other selections.
Null ‘split-sample’ search in IIS

Dummies included initially

Block 1

Dummies retained

Selected model: actual and fitted

actual

fitted

Block 2

Final

0.5

1.0

0

50

100

0.5

1.0

0

50

100

0.5

1.0

0

50

100

0.5

1.0

0

50

100

0.5

1.0

0

50

100

0.5

1.0

0

50

100

0.5

1.0

0

50

100

0.5

1.0

0

50

100

0.5

1.0

0

50

100

0.5

1.0

0

50

100

0.5

1.0

0

50

100

0.5

1.0

0

50

100

0.5

1.0

0

50

100
‘Split-sample’ search in IIS

Dummies included initially

Dummies retained

Selected model: actual and fitted

Block 1

Block 2

Final

David F. Hendry (INET at OMS)  Deciding between Alternative Approaches  ESRC/OMS Conference, 2012  17 / 49
Formulation decisions of which \( r \) variables \( w_t \); their lag lengths \( s \); cubics + exponentials in \( u_t \), after orthogonalizing \( w_t \); location shifts (any number, anywhere).

Leads to general unrestricted model (GUM):

\[
y_t = \sum_{i=1}^{r} \sum_{j=0}^{s} \beta_{i,j} w_{i,t-j} + \sum_{i=1}^{r} \sum_{j=0}^{s} \kappa_{i,j} u_{i,t-j}^2 + \sum_{i=1}^{r} \sum_{j=0}^{s} \theta_{i,j} u_{i,t-j}^3 + \sum_{i=1}^{r} \sum_{j=0}^{s} \gamma_{i,j} u_{i,t} e^{-|u_{i,t}|} + \sum_{j=1}^{s} \lambda_j y_{t-j} + \sum_{i=1}^{T} \delta_i 1\{i=t\} + \epsilon_t (3)
\]

\( K = 4r(s + 1) + s \) potential regressors, plus \( T \) indicators: bound to have \( N > T \)–consider exogeneity later.
Approach is **not** atheoretic.

**Theory formulations should be embedded in GUM, can be retained without selection, but does not guarantee they will be significant.**

Can also ensure theory-derived **signs** of long-run relation maintained, if not significantly rejected by the evidence.

**But much observed data variability in economics is due to features absent from most economic theories: which empirical models must handle.**

Extension of DGP candidates, $x_t$, in GUM allows theory formulation as special case, yet protects against contaminating influences (like outliers) absent from theory.

‘*Extras*’ can be selected at tight significance levels.
Correct \( n \) valid conditioning variables, \( z_t \), constant parameters \( \beta \):

\[
y_t = \beta' z_t + \epsilon_t
\]

where \( \epsilon_t \sim \text{IN}[0, \sigma^2_\epsilon] \), independently of \( z_t \). Then:

\[
\hat{\beta} = \beta + \left( \frac{1}{T} \sum_{t=1}^{T} z_t z_t' \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} z_t \epsilon_t \right) \sim N_n \left[ \beta, \sigma^2_\epsilon \left( \frac{1}{T} \sum_{t=1}^{T} z_t z_t' \right)^{-1} \right]
\]

Next, \( z_t \) retained during model selection over second set of \( k \) irrelevant candidate variables, \( w_t \), with coefficients \( \gamma = 0 \) when \( (k + n) << T \), so GUM is:

\[
y_t = \beta' z_t + \gamma' w_t + \nu_t
\]

Orthogonalize \( z_t \) and \( w_t \) by:

\[
w_t = \hat{\Gamma} z_t + u_t
\]

Then as \( \gamma = 0 \):

\[
y_t = \beta' z_t + \gamma' w_t + \nu_t = \beta' z_t + \gamma' u_t + \nu_t
\]

Coefficient of \( z_t \) unaltered.
Consequently:

\[
\begin{align*}
(\tilde{\beta} - \beta) &= \left( \sum_{t=1}^{T} z_t z'_t \quad \sum_{t=1}^{T} z_t u'_t \right) - 1 \left( \sum_{t=1}^{T} z_t v_t \right) \\
& \sim N_{n+k} \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2_e \begin{pmatrix} \left( \sum_{t=1}^{T} z_t z'_t \right)^{-1} \\ 0 \end{pmatrix} \begin{pmatrix} \left( \sum_{t=1}^{T} u_t u'_t \right)^{-1} \end{pmatrix} \right]
\end{align*}
\]

as \(\sum_{t=1}^{T} z_t u'_t = 0\), so distribution of \(\tilde{\beta}\) in (9) identical to that of \(\hat{\beta}\) in (5): unaffected by model selection.

Only costs of selection are:

(a) chance retentions of some \(u_t\) from selection, controlled by very tight significance levels \((\alpha \leq \min[0.01, 1/(N + T)]\); and

(b) impact on estimated distribution of \(\tilde{\beta}\) through \(\tilde{\sigma}^2_e\), offset by bias correction.
Two distinct forms of under-specification:
a] omitting relevant functions or lags of variables in DGP;
avoided by sufficiently general initial model:
b] omitting relevant variables, $\eta_t$, from the DGP;
induces less useful DGP—hard to avoid if $\eta_t$ unknown.

In a], $\gamma \neq 0$, as $z_t$ and $u_t$ orthogonal in (8), coefficient of former is $\beta + \gamma'\hat{\Gamma}$, which is estimated if (4) is simply fitted to the data: but may be significant with anticipated signs.

In b], when (6) nests DGP, but $\eta_t$ omitted from DGP, selection can substantively improve the final model:
see Castle and Hendry (2012).

**Win-win situation: theory kept if valid and complete; yet learn when it is not correct—empirical model discovery embedding theory evaluation.**

Can automatic model selection still work when $N > T$?
As many candidate variables as observations

Analytic approach to understanding IIS applies for \( N = T \) IID mutually orthogonal candidate regressors under the null.

Add first \( N/2 \) and select at significance level \( \alpha = 1/T = 1/N \). Record which were significant, and drop all.

Now add second block of \( N/2 \), again select at significance level \( \alpha = 1/N \), and record which are significant.

Finally, combine recorded variables from the two stages (if any), and select again at significance level \( \alpha = 1/N \).

At both sub-steps, on average \( \alpha N/2 = 1/2 \) a variable will be retained by chance, so on average \( \alpha N = 1 \) from the combined stage.

Again 99% efficient under the null at eliminating irrelevant variables—lose one degree of freedom on average.

If \( N > T \), divide in more sub-blocks; and if relevant variables to be retained, orthogonalize them with respect to the rest.
When breaks occur, estimated models constant only if:
(a) all substantive variables are included; and
(b) no internal shifts occur,
both demanding requirements.

Tackle (a) by commencing from large initial information set.

Super exogeneity violated by omitted variables:
policy model becomes non-constant when policy is changed.

Investigate in advance when previous policy changes:
examine marginal process for in-sample shifts;
test for co-breaking with policy model.

For (b), an indicator for location shifts would remove non-constancy,
as will IIS when it reflects such an indicator.

Location shifts induce changing error variances, so
\(|t| < c_\alpha\) if no indicators, yet larger with.
Parameter invariance essential in policy models: else mis-predict under regime shifts.

Automatic test of super exogeneity in Hendry and Santos (2010): IIS in marginal models, retain all significant outcomes and test their relevance in conditional model

No ex ante knowledge of timing or magnitudes of breaks: need not know DGP of marginal variables

Test has correct size under null of super exogeneity for a range of sizes of marginal-model saturation tests

Power to detect failures of super exogeneity when location shifts in marginal models: applies equally to models with expectations like NKPC: see Castle, Doornik, Hendry, and Nymoen (2012).
Non-invariance of NKPCs

‘Hybrid’ NKPC given by:

\[
\Delta p_t = \gamma_f E[\Delta p_{t+1} | J_t] + \gamma_b \Delta p_{t-1} + \pi s_t + u_t \geq 0
\]  \(10\)

\(\Delta p_t, s_t\) are rate of inflation & firms’ real marginal costs, so:

\[
\Delta p_t = \gamma_f \Delta p_{t+1} + \gamma_b \Delta p_{t-1} + \pi s_t + \epsilon_t, \quad \epsilon_t \sim D[0, \sigma^2_{\epsilon}]
\]  \(11\)

where:

\[
E[\Delta p_{t+1} | J_t] = \Delta p_{t+1} + \nu_{t+1}
\]  \(12\)

which appears to suggest that expectations are unbiased as:

\[
E[\nu_{t+1} | J_t] = 0.
\]  \(13\)

Then \(\Delta p_{t+1}\) instrumented by \(k\) variables \(z_t\) implicitly postulating:

\[
\Delta p_t = \kappa' z_t + \nu_t
\]  \(14\)

Assumes a constant-parameter world, so test by IIS on (14) by adding retained indicators to (11): significance refutes invariance.

Also, insignificance of \(\gamma_f\) inconsistent with forward-looking formulation.
Euro-area hybrid NKPC with IV estimation, $\Delta p_{t+1}$ and $s_t$ endogenous, using instruments: five lags of inflation, two lags of $s_t$, detrended output and wage inflation; sample $T = 102$ (1972(2) to 1998(1)):

$$
\hat{\Delta p}_t = 0.655 \hat{\Delta p}_{t+1} + 0.280 \Delta p_{t-1} + 0.012 s_t + 0.009 \\
\hat{\chi}^2_S(6) = 11.88
$$

(15)

Elasticities sum to 0.94 and $\hat{\gamma}_f$ comparable to reported GMM estimates: Galí and Gertler (1999).

Forecasting equation for $\Delta p_t$ uses instrument set for NKPC estimation with IIS in *Autometrics*: for $\alpha = 0.025$, finds 11 indicators.
Adding $gap_{t-1}$ and the 11 indicators makes NKPC congruent: $\chi^2_S(4) = 2.42$: no significant tests of residual mis-specification.

Nine of the 11 ‘reduced form’ indicators retained: clear evidence for lack of invariance in feed-forward NKPC.

**Coefficient of $\Delta p_{t+1}$ is negative and insignificantly different from zero.**

Coefficient of wage-share is sizeable, serves as an important equilibrating mechanism.

**Inflation ‘persistence’ an artifact of mis-specified NKPC model**

A failure to model breaks induces *spurious significance* of feed-forward terms proxying expectations: **but a deeper problem lurks.**
Write $f_{\Delta p_t}$ as density of $\Delta p_t$: then (13) is really $E_{f_{\Delta p_t}} [\nu_{t+1} | J_t] = 0$. Nothing ensures $E_{f_{\Delta p_{t+1}}} [\nu_{t+1} | J_t] = 0$ for unbiased expectations:

$$E_{f_{\Delta p_{t+1}}} [\Delta p_{t+1} | J_t] = \int \Delta p_{t+1} f_{\Delta p_{t+1}} (\Delta p_{t+1} | J_t) d\Delta p_{t+1}$$  \hspace{1cm} (16)

so (16) requires a crystal ball for future $f_{\Delta p_{t+1}} (\Delta p_{t+1} | J_t)$. 

Problem lies in expectations formation
Write $f_{\Delta p_t}$ as density of $\Delta p_t$: then (13) is really $E_{f_{\Delta p_t}} [\nu_{t+1} \mid J_t] = 0$. Nothing ensures $E_{f_{\Delta p_{t+1}}} [\nu_{t+1} \mid J_t] = 0$ for unbiased expectations:

$$E_{f_{\Delta p_{t+1}}} [\Delta p_{t+1} \mid J_t] = \int \Delta p_{t+1} f_{\Delta p_{t+1}} (\Delta p_{t+1} \mid J_t) \, d\Delta p_{t+1}$$

(16)

so (16) requires a crystal ball for future $f_{\Delta p_{t+1}} (\Delta p_{t+1} \mid J_t)$.

Best an agent can do is form a sensible expectation, forecasting $f_{\Delta p_{t+1}} (\cdot)$ by $\hat{f}_{\Delta p_{t+1}} (\cdot)$. If moments of $f_{\Delta p_{t+1}} (\cdot)$ alter, no good rules for $\hat{f}_{\Delta p_{t+1}} (\cdot)$, but $\hat{f}_{\Delta p_{t+1}} (\cdot) = f_{\Delta p_t} (\cdot)$ is not a good choice.

Agents cannot know $f_{\Delta p_{t+1}} (\cdot)$ if no time invariance: nor how $J_t$ enters.

Sleight of hand by not subscripting expectations by the relevant distribution: revealed by accounting for shifts.
Law of iterated expectations fails

When variables, say \((x_{t+1}, x_t)\), at different dates are drawn from same distribution \(f_x\):

\[
E_{f_x} [E_{f_x} [x_{t+1} | x_t]] = E_{f_x} [x_{t+1}] \tag{17}
\]

But when distributions shift:

\[
E_{f_{xt}} [E_{f_{xt+1}} [x_{t+1} | x_t]] \neq E_{f_{xt+1}} [x_{t+1}] \tag{18}
\]

as:

\[
f_{xt+1} (x_{t+1} | x_t) f_{xt} (x_t) \neq f_{xt+1} (x_{t+1} | x_t) f_{xt+1} (x_t) \tag{19}
\]

See Hendry and Mizon (2010) for formal derivations.

Invalidates inter-temporal derivations facing unanticipated shifts: or theory requires that no location shifts occur—so is empirically irrelevant.

DSGEs are intrinsically non-structural: mathematical basis fails when distributions alter.
Conclusions on empirical modelling

All essential steps feasible once target DGP defined:

1. automatically create general model from investigator’s $x_t$:
   extra variables, longer lags, non-linearity, & shift indicators—ensures congruent GUM and avoids huge costs from under-specified models;

2. embed theory-model, orthogonalizing other variables—ensures specification is retained unaltered;

3. select most parsimonious congruent encompassing model—ensures undominated representation;

4. compute near-unbiased parameter estimates—ensures appropriate quantification; and

5. stringently evaluate results, especially super exogeneity—ensures selected model valid, and usable.

Generalizes to $N > T$ with expanding and contracting searches.
Many risks from location shifts

- Forecast failure primarily due to location shifts;
- systematic mis-forecasting by all equilibrium-correction models: regressions, VARs, DSGEs, EqCMs, GARCH, etc.;
- difficult to predict, but can mitigate failure by robust devices;
- location shifts invalidate law of iterated expectations;
- conditional expectations not unbiased for next period;
- ‘rational expectations’ systematically biased after shifts;
- yet every DGP parameter can shift without noticeable effect;
- agents could not quickly learn what had changed;
- super exogeneity test can exploit location shifts;
- NKPC shows lack of invariance: insignificant feed-forward term and little inflation persistence.


Retracing route

(A) Empirical model discovery
(B) Multiple location shifts
(C) Evaluating theory models
(D) Evaluating policy models
(E) Non-invariance of NKPCs

Conclusion
Identification does not need prior information

When system is:

\[ y_t = \Psi z_t + v_t \]  \text{where}  \ v_t \sim \text{IN}_m [0, \Omega_v] \tag{20} 

with model:

\[ By_t = Cz_t + \epsilon_t \]  \text{where}  \ \epsilon_t \sim \text{IN}_m [0, \Sigma_\epsilon] \tag{21} 

uniqueness of \( B, C \) from solving \( B\Psi = C \) is intrinsic to DGP. If \( B, C \) identified, can be found without prior knowledge as (21) is a reduction of (20): see Hendry and Krolzig (2005). Confirmed by over-identified restrictions being testable by parsimonious encompassing.

If \( B, C \) not identified, (20) is least restricted identified representation; but all just-identified representations equivalent.

For given \( B, C \) over-identified in DGP, (21) is unique. But other \( B^*, C^* \) could be equally over-identified with equal likelihood.
Super exogeneity violated by omissions

To illustrate, consider conditional DGP:

$$ y_t = \beta_0 + \beta_1' x_{1,t} + \beta_2' x_{2,t} + \epsilon_t $$ (22)

where processes non-constant with:

$$ x_{2,t} = E[x_{2,t}] + \Gamma (x_{1,t} - E[x_{1,t}]) + v_t $$ (23)

$$ E[x_{1,t}] = \psi_{1,1} + \psi_{1,2} 1\{t \geq T_1\} \quad \text{and} \quad E[x_{2,t}] = \psi_{2,1} + \psi_{2,2} 1\{t \geq T_2\} $$ (24)

Not known $x_{2,t}$ is relevant, so DGP is:

$$ y_t = \beta_0 + \beta_2' (E[x_{2,t}] - \Gamma E[x_{1,t}]) + (\beta_1' + \beta_2' \Gamma) x_{1,t} + \beta_2' v_t + \epsilon_t $$ (25)

Intercept non-constant in (25) if $\Gamma \neq 0$.

Assumed model mis-specified by omitting $x_{2,t}$:

$$ y_t = \gamma_0 + \gamma_1' x_{1,t} + \epsilon_t $$ (26)

even if constant in-sample, when $x_{1,t}$ is changed to implement a policy, outcome will not be as anticipated.
From the DGP (22):

$$\frac{\partial y_t}{\partial x'_{1,t}} = \beta_1$$ \hspace{1cm} (27)

so if $x_{1,t}$ is shifted to $x_{1,t} + \nabla x_{1,t}$, then $y_t$ shifts by $\beta'_1 \nabla x_{1,t}$.

But (25) and (26) entail that shift is $(\beta'_1 + \beta'_2 \Gamma) \nabla x_{1,t}$, so intercept shift of $-\beta'_2 \Gamma \nabla x_{1,t}$ occurs in the model.

Policy model non-constant precisely when policy is changed: failure of super exogeneity.

Investigate in advance if any previous policy changes: examine marginal process of $x_{1,t}$ for in-sample shifts.

Ascertain occurrence and timing of location shifts in $x_{1,t}$, and check whether they coincide with shifts in conditional model $y_t | x_{1,t}$.
First stage is IIS in marginal, retaining dummies at significance level $\alpha_1$:

$$x_t = \pi_0 + \sum_{j=1}^{s} \Pi_j x_{t-j} + \sum_{i=1}^{m} \rho_{i,\alpha_1} 1\{t=t_i\} + v_{2,t}^*$$ \hspace{1cm} (28)

Second stage adds $m$ retained indicators to conditional:

$$y_t = \mu_0 + \beta' x_t + \sum_{i=1}^{m} \tau_{i,\alpha_2} 1\{t=t_i\} + \epsilon_t$$ \hspace{1cm} (29)

Conduct $F$-test for significance of $\left(\tau_1, \alpha_2 \ldots \tau_m, \alpha_2\right)$ at level $\alpha_2$

Test has power as significant impulse indicators capture outliers and location shifts not explained by regressors.
Basis of approach

Data generation process (DGP):
joint density of all variables in economy

Impossible to accurately theorize about or model precisely
Too high dimensional and far too non-stationary.

Need to reduce to manageable size in ‘local DGP’ (LDGP):
the DGP in space of \( r \) variables \( \{x_t\} \) being modelled

**Theory of reduction explains derivation of LDGP:**
joint density \( D_x(x_1 \ldots x_T|\theta) \).
Acts as DGP, but ‘parameter’ \( \theta \) may be time varying

Knowing LDGP, can generate ‘look alike data’ for \( \{x_t\} \)
which only deviate from actual data by unpredictable noise

Once \( \{x_t\} \) chosen, cannot do better than know \( D_x(\cdot) \)—
so the LDGP \( D_x(\cdot) \) is the *target* for model selection:
need to relate theory model to that target.
To establish ‘truth’ requires at least these 12 assumptions:

1. correct, comprehensive, & immutable economic theory;
2. correct, complete choice of all relevant variables & lags;
3. validity & relevance of all regressors & instruments;
4. precise functional forms for all variables;
5. absence of hidden dependencies;
6. all expectations formulations correct;
7. all parameters identified, constant over time, & invariant;
8. exact data measurements on every variable;
9. errors are ‘independent’ & homoscedastic;
10. error distributions constant over time;
11. appropriate estimator at relevant sample sizes;
12. valid and non-distortionary method of model selection.

If ‘truth’ is not on offer—what is?
Many features of models not derivable from theory.
Almost always must be data-based on available sample: need to **discover** what matters empirically.

Need empirical evidence on which:
[a]: variables are actually relevant (**specification**),
[b]: their lagged responses (**dynamic reactions**),
[c]: functional forms of relationships (**non-linearities**),
[d]: structural breaks & unit roots (**non-stationarities**),
[e]: simultaneity (or **exogeneity**), expectations, etc.

**Theory provides an object for modeling**—not the target:
(A) embed that object in much more **general formulation**;
(B) search for the **simplest acceptable representation**;
(C) **evaluate** the findings for congruence and encompassing.

**How to accomplish? And what are its properties?**
Implications for automatic methods

Seven stages for empirical model discovery and theory evaluation in econometrics.

First, theoretical derivation of $x$, and putative relationships.

Second, **automatically create more general model** in $w$ embedding $y = f(x)$.

Third, **automatic selection** over orthogonalized representations, retaining theory.

Fourth, end with **congruent parsimonious-encompassing model**.

Fifth, **quantify** the outcome by **unbiasedly estimating resulting model**.

Sixth, **evaluate** theory directly and any discoveries on **new data, new tests and new procedures**.

Seventh, **summarize** possibly vast information set in **parsimonious but undominated model**.
Selecting by *Autometrics* with $\alpha = 0.05$:

\[
\hat{\Delta p}_t = -0.298 \hat{\Delta p}_{t+1} + 0.115 s_t + 0.505 \Delta p_{t-1} + 0.086 + 0.0015 \text{ gap}_{t-1} \\
+ 1.10 I_{73}(1),t + 1.09 I_{73}(3),t + 0.73 I_{73}(4),t + 0.85 I_{74}(2),t \\
+ 0.80 I_{74}(3),t + 0.98 I_{76}(2),t + 0.57 I_{76}(3),t - 0.65 I_{78}(4),t + 0.69 I_{83}(1),t
\]

\[\chi^2_S(6) = 5.06 \quad F_{ar}(5, 85) = 1.55 \quad F_{arch}(4, 96) = 1.49 \]
\[F_{het}(14, 80) = 1.41 \quad \chi^2_{nd}(2) = 1.04 \]

(coefficients of dummies multiplied by 100)

$F_{name}$ denotes an approximate $F$-test:

$F_{ar}$ for $k^{th}$-order serial correlation;

$F_{het}$ for heteroskedasticity;

$F_{reset}$ for functional form;

$F_{arch}$ for $k^{th}$-order ARCH; and

$\chi^2_{nd}(2)$ for normality.
Problems with forecasting

Consider stationary scalar AR(1) DGP with known exogenous 
\( \{ z_t \} \sim \text{IN}[0, 1] \):

\[
x_t = \mu + \rho x_{t-1} + \gamma z_t + \epsilon_t \quad \text{where} \quad \epsilon_t \sim \text{IN}[0, \sigma^2_{\epsilon}] \quad \text{and} \quad |\rho| < 1 \quad (30)
\]

over \( t = 1, \ldots, T \) where \( E[z_t] = \kappa \).

When \( \mu, \rho, \) and \( \gamma \) known & constant, forecast from \( x_T \) for known \( z_{T+1} \):

\[
\hat{x}_{T+1|T} = \mu + \rho x_T + \gamma z_{T+1} \quad (31)
\]

is unbiased for \( x_{T+1} \) with smallest possible variance.

What parameter shifts cause problems?

Location shifts alone are problematic, namely shifts in \( \theta = (\mu + \gamma \kappa)/(1 - \rho) \):

- change all parameters, forecasting with mis-specified model—\textbf{no failure};
- change just \( \rho \) using in-sample DGP with zero intercept—\textbf{massive failure}. 

David F. Hendry (INET at OMS)  Deciding between Alternative Approaches  ESRC/OMS Conference, 2012
Incorrect specification, \( \mu = \kappa = 10 \) and \( \rho \) changed twice

Model incorrectly specified, \( z_{T+1} \) omitted with \( \kappa = 10 \). Forecasts after breaks in \( \rho = 0.8 \) to \( \rho^* = 0.4 \), & \( \mu = 10 \) to \( \mu^* = 50 \) at \( T = 41 \) then back at \( T = 46 \) so:

\[
\chi_{T+h} = \mu^* + \rho^* \chi_{T+h-1} + \gamma z_{T+h} + \epsilon_{T+h}
\]

(32)

Yet no forecast failure when \( \hat{\chi}_{T+h|T+h-1} = \hat{\mu} + \hat{\rho} \chi_{T+h-1} \).
Model correctly specified in-sample, forecasts for same break, $\mu = 0$, and $\kappa = 10$ but $z_{t+1}$ included. Forecast failure is manifest. In-sample correct specification need not help even with a zero intercept and known future $z_{T+h}$. 
Many parameters shift

\begin{align*}
\mu &= 0; \gamma = 2; \kappa = 5; \rho = 0.8; \\
\mu &= 0; \gamma^* = 1.36; \kappa = 5; \rho^* = 0.6
\end{align*}

\begin{align*}
\mu &= 5; \gamma = 1; \kappa = 5; \rho = 0.8; \\
\mu^* &= 2.5; \gamma^* = 0.86; \kappa = 5; \rho^* = 0.6
\end{align*}

\begin{align*}
\mu &= 5.0; \gamma = 0.25; \kappa = 20; \rho = 0.8 \\
\mu^* &= 0.5; \gamma^* = 0.4; \kappa = 20; \rho^* = 0.5
\end{align*}

\begin{align*}
\mu &= 5; \gamma = 2.5; \kappa = 2; \rho = 0.8; \\
\mu^* &= -2.95; \gamma^* = 7; \kappa = 2; \rho^* = 0.35
\end{align*}

Can essentially replicate break by changing \( \mu, \gamma, \kappa \) and \( \rho \) in many combinations: economic agents could not tell what had shifted till long afterwards. Analysis in Gabaix (2012) is infeasible.
Robust forecasts avoid systematic failure.

Difference mis-specified model:

\[
\Delta \tilde{x}_{T+h|T+h-1} = \hat{\rho} \Delta x_{T+h-1} \quad \text{or:} \quad \tilde{x}_{T+h|T+h-1} = x_{T+h-1} + \hat{\rho} \Delta x_{T+h-1}
\]

Uses ‘wrong’ \( \hat{\rho} \) for first 5 forecasts;
incorrectly differenced;
and omits relevant variable.

But no location shifts.
Robust forecasting device (33) **avoids most of last 9 forecast errors.** RMS FE of estimated DGP (30) is 6.6 versus 5.5 here; but 3.8 versus 2.0 over last 9 forecasts, so nearly halved.

**Never judge a model’s verisimilitude by forecast performance.**