Empirical economic model discovery and theory evaluation
David Hendry, Oxford, United Kingdom

ABSTRACT:
Economies are too high dimensional and wide sense non-stationary for all features of models to be derived by either prior reasoning or data modelling alone. Selecting a viable representation intrinsically involves empirical discovery jointly with theory evaluation. Automatic methods can formulate very general initial specifications with many candidate variables, long lag lengths, and non-linearities, while allowing for outliers and location shifts at every observation, then select congruent parsimonious-encompassing models. Theory-relevant variables are retained without selection, while selecting other candidate variables. Under the null that the latter are irrelevant, by orthogonalizing with respect to the theory variables, estimator distributions of the theory-model’s parameters are unaffected by selection, even for more variables than observations and for endogenous variables. Under the alternative, when the initial model nests the local data generating process, an improved outcome results from selection, allowing rigorous evaluation of any postulated models to ascertain their validity.
EMPIRICAL MODEL DISCOVERY AND THEORY EVALUATION

David F. Hendry
Program in Economic Modelling, INET Oxford
March 2014, OxMetrics Conference, Washington

Research jointly with Jennifer Castle, Jurgen Doornik and Søren Johansen
Every decision about:

1. a theory formulation;
2. its implementation;
3. its evidential base;
4. its empirical specification; and
5. its evaluation

involves selection.

Absent omniscience, selection is inevitable, unavoidable and ubiquitous: issue is not whether to select, but how to select.
Omniscience is not on offer

Data generation process (DGP)

Economic mechanism plus measurement system. Economies are high dimensional, interdependent, heterogeneous, and evolving: a comprehensive specification of all events is impossible. Aggregation over time, space, commodities, agents, endowments, is essential—but preclude claims to ‘truth’.

Local DGP (LDGP) is DGP for $n$ variables $\{x_t\}$ under analysis:

joint density $D_x(x_1 \ldots x_T | \theta)$.

Acts as DGP, but ‘parameter’ $\theta$ may be time varying.

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.
Empirical models reflect LDGP, not facsimiles: designed to satisfy—often implicit—selection criteria.

Only congruence is on offer in economics: congruent models match LDGP in all measured attributes. ‘True’ models in class of congruent models. Congruence is testable: necessary conditions for structure.

**Encompassing**: explain the results of all other models.

Theory only provides an object for modelling:
(A) embed that object in the initial general formulation;
(B) search for the simplest acceptable representation;
(C) evaluate the findings.

How to accomplish that? And what are its properties?
Empirical model formulation

Seven categories of evidence matter jointly

(i) many candidate explanatory variables;
(ii) dynamic reactions;
(iii) parameter changes and location shifts;
(iv) relationships may be non-linear;
(v) feedbacks, exogeneity, and expectations;
(vi) evaluating congruence;
(vii) encompassing results of rival models.

To successfully determine what matters and how it enters, all potential determinants must be included: omitting key variables adversely affects selected models.

As macroeconomic variables are highly intercorrelated, initially need large equations to capture all these effects.
Catch 22—and its resolution

Especially forceful issue when data processes are ‘wide sense non-stationary’: integrated and not time invariant.

Often leads to more variables $N$ than observations $T$.

‘Catch 22’—if $N > T$, everything cannot be entered from the outset: necessitates iterative search algorithms to eliminate irrelevant.

To resolve conundrum, analysis proceeds in nine stages.

[C] Doornik (2009), Autometrics.
[H] Hendry and Santos (2010), Testing super exogeneity.
Nine stages

1] ‘1-cut’ selection for orthogonal designs with \( N << T \); establishes ‘good behaviour’ of selection \textit{per se}: [A].

2] Selection matters, so derive \textbf{bias corrections} for conditional distributions; improves mean-square errors (MSEs): [B].

3] Compare ‘1-cut’ with \textit{Autometrics} (applicable to non-orthogonal models); shows \textit{Autometrics} outperforms, & can handle \( N > T \): [C].

4] \textbf{Indicator saturation} for multiple shifts and outliers; now \( N > T \) must occur: [D].

5] Selecting non-linearities: [E].

6] Impact of \textbf{mis-specification testing}; costs of checking congruence small compared to not testing: [A].

7] Role of \textbf{encompassing} in automatic selection; controls ‘good behavior’ & avoids missing relevant combinations: [G].

8] Empirical model discovery \textit{jointly with theory evaluation}: [F].

9] Finally, testing \textbf{exogeneity} in selected model: [H].
(1) Selecting empirical models
(2) Simulating ‘1-cut’ selection
(3) Automatic model extensions: *Autometrics*
(4) Detecting and modelling multiple location shifts
(5) Mis-specification testing and encompassing
(6) Empirical model discovery and theory evaluation
(7) Modelling UK real wages over the last 150 years
(8) Conclusions
Aim for final selection that maintains congruence of GUM, and parsimoniously encompasses it, so is ‘best’ representation of LDGP. Embodied in PcGive & Autometrics: see Doornik and Hendry (2013).
Aim for frequency of recovering LDGP starting from GUM same as starting from LDGP.

Two costs of selection: costs of inference and costs of search.

First inevitable if tests have non-zero null retention and non-unit rejection frequencies under alternative: applies even if commence from LDGP.

Avoid for theory parameters by embedding theory without search.

Measure costs of inference by RMSE of selecting or conducting inference on LDGP.

When a GUM nests the LDGP, additional costs of search: calculate by increase in RMSEs for relevant variables when starting from GUM as against LDGP, plus costs for retained irrelevant variables.
(1) Selecting empirical models
(2) Simulating ‘1-cut’ selection
(3) Automatic model extensions: *Autometrics*
(4) Detecting and modelling multiple location shifts
(5) Mis-specification testing and encompassing
(6) Empirical model discovery and theory evaluation
(7) Modelling UK real wages over the last 150 years
(8) Conclusions
Gauge and potency

‘gauge’ is null retention frequency of selection statistics.

‘potency’ is average non-null retention frequency.

\( \hat{\beta}_{k,i} \) is OLS estimate of coefficient on \( x_{k,t} \) in GUM for replication \( i \).

\( \tilde{\beta}_{k,i} \) is OLS estimate after selection ( \( \tilde{\beta}_{k,i} = 0 \) if \( z_{k,t} \) not selected).

retention rate:

\[
\tilde{p}_k = \frac{1}{M} \sum_{i=1}^{M} 1(\hat{\beta}_{k,i} \neq 0), \quad k = 0, \ldots, N,
\]

potency:

\[
\text{potency} = \frac{1}{n} \sum_{k=1}^{n} \tilde{p}_k,
\]

gauge:

\[
\text{gauge} = \frac{1}{N-n+1} \left( \tilde{p}_0 + \sum_{k=n+1}^{N} \tilde{p}_k \right).
\]

CMSE is conditional MSE:

\[
\text{CMSE}_k = \sum_{i=1}^{M} \frac{\left( \hat{\beta}_{k,i} - \beta_k \right)^2 \cdot 1(\hat{\beta}_{k,i} \neq 0)}{\sum_{i=1}^{M} 1(\hat{\beta}_{k,i} \neq 0)}, \quad \left( \beta_k^2 \text{ if } \sum_{i=1}^{M} 1(\hat{\beta}_{k,i} \neq 0) = 0 \right)
\]

GUM includes all \( N \) variables (1001 here with intercept):

\[
y_t = \beta_0 + \beta_1 z_{1,t} + \cdots + \beta_{1000} z_{1000,t} + \nu_t
\] (1)
Simulation outcomes

DGP is given by:

\[ y_t = \beta_1 z_{1,t} + \cdots + \beta_{10} z_{10,t} + \epsilon_t, \quad (2) \]
\[ z_t \sim \text{IN}_{1000} [0, \Omega], \quad (3) \]
\[ \epsilon_t \sim \text{IN} [0, 1], \quad (4) \]

where \( z'_t = (z_{1,t}, \cdots, z_{1000,t}) \), \( \Omega = I_{1000} \), \( T = 2000 \): non-centralities of \( \beta_i \) are \( \psi_i = 1.5 + 0.5i \) (so 2,...,6.5).

Table: Potency and gauge for 1-cut selection with \( N = 1000 \) variables.

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>gauge</th>
<th>potency</th>
<th>theory</th>
<th>power</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>1.01%</td>
<td>81%</td>
<td>81%</td>
<td></td>
</tr>
<tr>
<td>0.1%</td>
<td>0.10%</td>
<td>69%</td>
<td>68%</td>
<td></td>
</tr>
</tbody>
</table>

Gauges not significantly different from nominal sizes \( \alpha \): selection is not ‘oversized’ even with 1000 variables.

Potencies close to average theory powers of 0.811 and 0.684.

Close match between theory and evidence even when selecting just 10 relevant regressors from 1000 variables.
Retention rates for relevant variables match theory, yet model reduced by about 990 variables on average. Bias corrections when $|t| \geq c_\alpha$ improve further.
Remarkable decrease in MSEs of retained irrelevant variables when bias correction—despite not knowing which are irrelevant and which relevant variables. For $N = 1000$ and $n = 10$ in (??):

Table: Average CMSEs, times 100, for retained relevant and irrelevant variables (excluding $\beta_0$), with and without bias correction.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>1%</th>
<th>0.1%</th>
<th>1%</th>
<th>0.1%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>average CMSE over 990 irrelevant variables</td>
<td>average CMSE over 10 relevant variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>uncorrected $\tilde{\beta}$</td>
<td>0.84</td>
<td>1.23</td>
<td>1.0</td>
<td>1.4</td>
</tr>
<tr>
<td>$\tilde{\beta}$ after correction</td>
<td>0.38</td>
<td>0.60</td>
<td>1.2</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Greatly reduces MSEs of irrelevant variables in both unconditional and conditional distributions.

Coefficients of retained variables with $|t| \leq c_\alpha$ are not bias corrected—insignificant estimates set to zero.
Bias correcting conditional distributions at 5%

(a) $\psi = 2$

(b) $\psi = 4$

(c) $\psi = 0$

(d) Intercept, $\psi = 0$
(1) Selecting empirical models
(2) Simulating ‘1-cut’ selection
(3) **Automatic model extensions**: *Autometrics*
(4) Detecting and modelling multiple location shifts
(5) Mis-specification testing and encompassing
(6) Empirical model discovery and theory evaluation
(7) Modelling UK real wages over the last 150 years
(8) Conclusions
Extensions determine how well LDGP is approximated

Create three extensions automatically:

(i) lag formulation to implement **sequential factorization**;
(ii) functional form transformations for **non-linearity**;
(iii) indicator saturation (IIS/SIS) for **parameter non-constancy**.

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

$$y_t = \beta_0 + \sum_{i=1}^{s} \lambda_i y_{t-i} + \sum_{i=1}^{r} \sum_{j=0}^{s} \beta_{i,j} z_{i,t-j} + \epsilon_t$$  \hspace{1cm} (5)

$x_t = (y_t, z_t)$ could also be modelled as a system:

$$x_t = \gamma + \sum_{j=1}^{s} \Gamma_j x_{t-j} + \epsilon_t$$  \hspace{1cm} (6)

We focus on single equations, but systems can be handled.
(ii) Approximate non-linearity by functions of principal components $w_t$ of the $z_t$: Castle and Hendry (2010).

Let $z_t \sim D_n [\mu, \Omega]$, where $\Omega = H\Lambda H'$ with $H'H = I_n$.

Then $w_t^* = H'z_t \Rightarrow w_t^* \sim D_n [H'\mu, \Lambda]$.

Empirically $\hat{\Omega} = T^{-1} \sum_{t=1}^{T} (z_t - \bar{z})(z_t - \bar{z})' = \hat{H}\hat{\Lambda}\hat{H}'$ so that $w_t = \hat{H}'(z_t - \bar{z})$.

Implemented by squares, cubics and exponential functions:

$u_{1,i,t} = w_{i,t}^2$; $u_{2,i,t} = w_{i,t}^3$; $u_{3,i,t} = w_{i,t} e^{-|w_{i,t}|}$.

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

Number of potential regressors for cubic polynomials is:

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

Explosion in number of terms as $K$ increases:

<table>
<thead>
<tr>
<th>$K$</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_K$</td>
<td>3</td>
<td>55</td>
<td>285</td>
<td>1539</td>
<td>5455</td>
<td>12300</td>
</tr>
</tbody>
</table>

Quickly reach huge $M_K$: but only $3K$ if use $u_{k,i,t}$, $k = 1, 2, 3$. 

David F. Hendry (INET Oxford)  Empirical Model Discovery  November 2013 19 / 49
(1) Selecting empirical models
(2) Simulating ‘1-cut’ selection
(3) Automatic model extensions: Autometrics
(4) Detecting and modelling multiple location shifts
(5) Mis-specification testing and encompassing
(6) Empirical model discovery and theory evaluation
(7) Modelling UK real wages over the last 150 years
(8) Conclusions
Detecting multiple shifts

‘Portmanteau’ approach to detect location shifts anywhere in sample while also selecting over many candidate variables, lags etc.

Impulse-indicator saturation

IIS creates complete set of indicator variables:

\[
\{1_{\{j=t\}}\} = 1 \text{ when } j = t, \text{ and } 0 \text{ otherwise for } j = 1, \ldots, T.
\]

Add all \(T\) indicators to set of candidate variables when \(T\) observations.

Feasible ‘split-sample’ algorithm:


Include first half of indicators, record significant on 1-cut: ‘dummying out’ first \(T/2\) observations when estimating parameters.

Omit first half of indicators, include other half, record again.

Combine retained sub-sample indicators, & select significant.

\(\alpha T\) indicators selected on average at significance level \(\alpha\).

Chow (1960) test is sub-sample IIS over \(T - k + 1\) to \(T\).

Salkever (1976) tests parameter constancy by impulse indicators.
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 $\Sigma$:

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

Efficiency of IIS estimator $\tilde{\beta}$ with respect to OLS $\hat{\beta}$ measured by $\Omega_\alpha$ depends on $c_\alpha$ and distribution, but close to $(1 - \alpha)^{-1}I_n$.

Must lose efficiency under null; small loss $\alpha T$ of 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: but can be done jointly with all other selections.
Add a complete set of step indicators $S_1 = \{1_{t \leq j}, j = 1, \ldots, T\}$, where $1_{t \leq j} = 1$ for observations up to $j$, and zero otherwise. Step indicators cumulate impulse indicators up to each next observation.

**IIS: Impulses**

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

**SIS: Steps**

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

SIS has correct null retention frequency in constant conditional models for a nominal test size of $\alpha$, and a higher probability than IIS of finding location shifts.
Illustrating ‘split-half’ SIS for a single location shift

Add half indicators and select ones significant at 1\%.

Indicators included initially

Indicators retained

Selected model: actual and fitted

David F. Hendry  (INET Oxford)  Empirical Model Discovery  November 2013  24 / 49
Illustrating ‘split-half’ SIS for a single location shift

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

Indicators included initially

Indicators retained

Selected model: actual and fitted

![Graphs showing indicators and selected models.](image-url)
Illustrating ‘split-half’ SIS for a single location shift

Combine retained indicators and re-select at \( 1\% \).

Initially retains last step as mean shifts down, then finds location shift, so eliminates redundant indicator: just one step needed.
Specification of GUM

Formulation decisions of which \( r \) variables \( z_t \); their maximum lag lengths \((s)\); squares, cubics + exponentials in \( w_t \), after orthogonalizing \( z_t \); location shifts (any number, anywhere) by IIS and/or SIS.

Leads to general unrestricted model (GUM):

\[
y_t = \sum_{i=1}^{r} \sum_{j=0}^{s} \beta_{i,j} z_{i,t-j} + \sum_{i=1}^{r} \sum_{j=0}^{s} \kappa_{i,j} w_{i,t-j}^2 + \sum_{i=1}^{r} \sum_{j=0}^{s} \psi_{i,j} w_{i,t-j}^3 \\
+ \sum_{i=1}^{r} \sum_{j=0}^{s} \gamma_{i,j} w_{i,t-j} e^{-|w_{i,t-j}|} + \sum_{j=1}^{s} \lambda_j y_{t-j} \\
+ \sum_{i=1}^{T} \delta_{i} 1\{i=t\} + \sum_{i=1}^{T-1} \phi_{i} 1\{i \leq t\} + \epsilon_t \tag{7}
\]

\( K = 4r(s + 1) + s + T \) potential regressors (possibly \((2T - 1)\) indicators): bound to have \( N > T \)—exogeneity considered later.
(1) Selecting empirical models
(2) Simulating ‘1-cut’ selection
(3) Automatic model extensions: *Autometrics*
(4) Detecting and modelling multiple location shifts
(5) **Mis-specification testing and encompassing**
(6) Empirical model discovery and theory evaluation
(7) Modelling UK real wages over the last 150 years
(8) Conclusions
Selection effects on tests

Little impact of selection on test statistics. Small change in quantiles above nominal significance level: but increasing impact as quantile decreases. Bound to occur: models with significant heteroscedasticity not selected.

Not a ‘distortion’ of sampling properties: decision is taken for GUM. Conditional on that, no change should occur.

Next Figure reports QQ plots of actual against reference distributions under the null for the main mis-specification tests in DGP, GUM and selected model.
Selection effects on test distributions

DGP: QQ Plots of Diagnostics for M=1000 and T=100

GUM: QQ Plots of Diagnostics for M=1000 and T=100

Specific: QQ Plots of Diagnostics for M=1000 and T=100
*Autometrics* conducts inferences for I(0)
Most selection tests remain valid:
see Sims, Stock, and Watson (1990)
**Only tests for a unit root need non-standard critical values**

Implementing system cointegration in *Autometrics*

Most diagnostic tests also valid for integrated series:
see Wooldridge (1999)

**Heteroscedasticity tests an exception:**
powers of variables then behave oddly
see Caceres (2007)
Encompassing plays a key role

Variables removed only when new model is a valid reduction of GUM. Reduction fails if selection does not parsimoniously encompass GUM at $c_\alpha$: see Hendry (1995), §14.6.

If fails, variable retained despite insignificance on $t$-test, as in Doornik (2008).

*Autometrics* without encompassing loses both gauge and potency.

*Autometrics* with encompassing is well behaved:
gauge is close to nominal rejection frequency $\alpha$.
potency is close to theory maximum of 1-off $t$-test.
(1) Selecting empirical models
(2) Simulating ‘1-cut’ selection
(3) Automatic model extensions: *Autometrics*
(4) Detecting and modelling multiple location shifts
(5) Mis-specification testing and encompassing
(6) **Empirical model discovery and theory evaluation**
(7) Modelling UK real wages over the last 150 years
(8) Conclusions
Approach is not atheoretic.

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

Embed initial economic analysis \( y = f(z) \) in GUM, to be retained without selection, but does not guarantee parameters will be significant.

Extension of LDGP 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.

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

General approach explained in Castle, Doornik, and Hendry (2011).
Correct $n$ valid conditioning variables, $z_t$, constant parameters $\beta$:

$$y_t = \beta' z_t + \epsilon_t$$  \hspace{1cm} (8)

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

$$\hat{\beta} = \beta + \left( \sum_{t=1}^{T} z_t z_t' \right)^{-1} \sum_{t=1}^{T} z_t \epsilon_t \sim \mathcal{N}_n \left[ \beta, \sigma^2_{\epsilon} \left( \sum_{t=1}^{T} z_t z_t' \right)^{-1} \right]$$  \hspace{1cm} (9)

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:

$$y_t = \beta' z_t + \gamma' w_t + \nu_t$$  \hspace{1cm} (10)

Orthogonalize $z_t$ and $w_t$ by:

$$w_t = \hat{\Gamma} z_t + u_t$$  \hspace{1cm} (11)

Then as $\gamma = 0$:

$$y_t = \beta' z_t + \gamma' w_t + \nu_t = \beta' z_t + \gamma' u_t + \nu_t$$  \hspace{1cm} (12)

Coefficient of $z_t$ unaltered.
Consequently:

\[
\left( \tilde{\beta} - \beta \right) = \left( \sum_{t=1}^{T} z_t z_t' \sum_{t=1}^{T} u_t u_t' \right)^{-1} \left( \sum_{t=1}^{T} z_t \gamma_t \right)
\]

\[
\sim N_{n+k} \begin{pmatrix}
0 \\
0
\end{pmatrix}, \sigma^2_{\epsilon} \begin{pmatrix}
\left( \sum_{t=1}^{T} z_t z_t' \right)^{-1} & 0 \\
0 & \left( \sum_{t=1}^{T} u_t u_t' \right)^{-1}
\end{pmatrix}
\]

(13)

as \( \sum_{t=1}^{T} z_t u_t' = 0 \), so distribution of \( \tilde{\beta} \) in (13) **identical** to that of \( \hat{\beta} \) in (9), **unaffected** by model selection.

Only costs of selection are:

(a) chance retentions of some \( u_t \) from selection, controlled by very tight significance levels \( \alpha \leq \min[0.001, 1/(N + T)] \); and

(b) impact on **estimated** distribution of \( \tilde{\beta} \) through \( \tilde{\sigma}^2_{\epsilon} \), offset by bias correcting.
More candidate variables than observations

If also have relevant variables to be retained, and $N > T$, orthogonalize them with respect to the rest.

As $N > T$, divide in more sub-blocks, setting $\alpha = 1/N$.

Model retains desired sub-set of $n$ variables at every stage, and only selects over putative irrelevant variables at stringent significance level: under the null, has no impact on estimated coefficients of relevant variables, or their distributions.

Almost costless to check large numbers of candidate variables: huge benefits if initial specification incorrect, but enlarged GUM nests LDGP.

Have answers to every ‘seminar question’ before they are asked!
**Hoover–Perez experiments**

\[ T = 139, \textbf{3 relevant and 37 irrelevant variables: all \%} \]

<table>
<thead>
<tr>
<th></th>
<th>HP7</th>
<th>HP8</th>
<th>step-wise HP7</th>
<th>HP8</th>
<th>Lasso: BIC HP7</th>
<th>HP8</th>
<th>Autometrics HP7</th>
<th>HP8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauge</td>
<td>3.0*</td>
<td>0.9*</td>
<td>0.9</td>
<td>3.1</td>
<td>19.5</td>
<td>35.1</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td>Potency</td>
<td>94.0</td>
<td>99.9</td>
<td>100.0</td>
<td>53.3</td>
<td>94.4</td>
<td>86.3</td>
<td>99.2</td>
<td>100.0</td>
</tr>
<tr>
<td>DGP found</td>
<td>24.6</td>
<td>78.0</td>
<td>71.6</td>
<td>22.0</td>
<td>0.1</td>
<td>0.0</td>
<td>68.3</td>
<td>68.8</td>
</tr>
</tbody>
</table>

1% nominal size

* Only counting significant terms (tiebreaker was best-fitting model)

\[ T = 139, \textbf{3 relevant and 141 irrelevant variables} \]

<table>
<thead>
<tr>
<th></th>
<th>step-wise HP7</th>
<th>HP8</th>
<th>Autometrics HP7</th>
<th>HP8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauge</td>
<td>0.1</td>
<td>0.7</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>Potency</td>
<td>99.7</td>
<td>40.3</td>
<td>97.4</td>
<td>100.0</td>
</tr>
<tr>
<td>DGP found</td>
<td>87.4</td>
<td>9.0</td>
<td>82.9</td>
<td>90.2</td>
</tr>
</tbody>
</table>

0.1% nominal size

\[ \text{Large increase in probability of locating DGP relative to } \alpha = 0.01 \]

\[ \text{–so should not select by \textquoteleft\textquoteleft goodness of fit\textquoteright\textquoteright} \]
Approximating small relevant effects by PCs

To capture potential omissions of individually insignificant relevant effects, add $w_t$, or principal components, $w_{1,t}$, of unselected $z_t$. Could also reflect common trends modelled by latent factors. Effective when factor structure of $z_t$ matches relation between $y_t$ and $z_t$ in LDGP: then by representing individually-insignificant effects in $z_t$ by $w_{1,t}$, can achieve substantive reductions in RMSEs relative to just estimating the LDGP.

<table>
<thead>
<tr>
<th>$\alpha = 0.01$</th>
<th>[A]</th>
<th>[B]</th>
<th>[C]</th>
<th>[D]</th>
<th>[E]</th>
<th>[F]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 10$; $\psi_i = 1$; $\rho = 0.9$, $\sigma = 1$</td>
<td>mean $\hat{\sigma}$</td>
<td>1.00</td>
<td>1.01</td>
<td>1.01</td>
<td>1.05</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>mean Bias</td>
<td>-0.02</td>
<td>72.0</td>
<td>61.0</td>
<td>-0.02</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>mean RMSE</td>
<td>0.32</td>
<td>0.75</td>
<td>0.68</td>
<td>0.32</td>
<td>0.08</td>
</tr>
</tbody>
</table>

(A) estimating DGP; (B) selection from DGP by *Autometrics*; (C) bias correction of (B); (D) estimation of factor model; (E) 1-cut selection from factor model; (F) bias correction of (E).
(1) Selecting empirical models
(2) Simulating ‘1-cut’ selection
(3) Automatic model extensions: *Autometrics*
(4) Detecting and modelling multiple location shifts
(5) Mis-specification testing and encompassing
(6) Empirical model discovery and theory evaluation
(7) **Modelling UK real wages over the last 150 years**
(8) Conclusions
Example of empirical model discovery in action.

1. Examine roles of many regressors, dynamics, non-linearities, and shifts for integrated data (nominal wages rose by 68,000%).
2. Important wage-price spiral interactions.
3. Non-linear unemployment reaction.
4. Location shifts and outliers tackled by SIS.
5. Test exogeneity of all contemporaneous regressors.
6. Extended data set to forecast real wages over ‘Great Recession’.

All aspects must be modelled jointly for a coherent economic explanation, including all substantively relevant variables, their dynamics, shifts, and non-linearities.
Wages, prices, productivity and unemployment

\[ w \]

\[ p \]

\[ (y-l) \]

\[ (w-p) \]

\[ \Delta w \]

\[ \Delta p \]

\[ \Delta (w-p) \]

\[ (w-p-y+l) \]

\[ U_r \]
SIS and regime shifts

\[ \Delta(w-p)_t \]

SIS outcome

\[ \Delta p_t \]

SIS outcome

\[ \Delta(y-l)_t \]

SIS outcome

\[ U_{r,t} \]

SIS outcome

SIS reveals location shifts unconditionally:
two major shifts in \( \Delta(w-p)_t \) and \( \Delta(y-l)_t \), but different magnitudes at different times; huge outliers do not align.

Location shifts in \( U_{r,t} \) and \( \Delta p_t \) also do not match.
Non-linearity chosen functions

Non-linearity test significant at $p = 0.006$ with $F(36, 91) = 1.95$.

Two elements stood out:

**Non-linear real-wage reaction to inflation represented by:**

$$f_t \Delta p_t = \frac{-1}{1 + 1000(\Delta p_t)^2} \Delta p_t.$$

$(U_{rt} - 0.05)^2$ was an important additional non-linearity.

Selection at $\alpha = 0.001$ for the step indicators, retaining all economic variables (see Hendry and Johansen, 2014), then selected over those at $\alpha = 0.01$.

No diagnostic tests significant with $\hat{\sigma} = 1.04\%$ and $\text{RMSFE}=1.05\%$ over 2005–2011.
Final model

Final selection

\[
\Delta (w - p)_t = 0.021 + 0.35 \Delta (y - l)_t + 0.12 \Delta^2 (y - l)_{t-1} - 0.13 \Delta^2 p_{t-1} \\
- 0.18 (w - p - y + l - \mu)_{t-2} - 0.18 (U_{r,t} - 0.05) \\
+ 2.7 (U_{r,t} - 0.05)^2 - 0.13 \Delta^2 U_{r,t} + 0.71 (f_t \Delta p_t) - 0.15 S_{1939} \\
+ 0.18 S_{1940} - 0.06 S_{1941} - 0.024(S_{2011} - S_{1946}) \Delta u_{r,t} \\
- 0.0361_{1916} + 0.027 (1_{1942} + 1_{1943} - 1_{1944} - 1_{1945}) - 0.0441_{1977}
\]

\[
R^2 = 0.82; \quad \hat{\sigma} = 1.04\%; \quad SIC = -5.85; \quad T = 1864 - 2004;
\]

\[
\chi^2_{nd}(2) = 2.26; \quad F_{ar}(2, 123) = 0.39; \quad F_{arch}(1, 139) = 0.49;
\]
\[
F_{het}(20, 116) = 0.82; \quad F_{reset}(2, 124) = 2.28; \quad F_{chow}(7, 125) = 0.95.
\]

\[u_{r,t} = \log(U_{r,t})\] and \(\hat{\mu}\) is the sample mean of \((w - p - y + l)\).

(e.g.) \(S_{1939}\) is step indicator: 1 till 1939 and 0 after, etc.
Short-run impact of changes in productivity is $\approx 0.6$

Strong equilibrium correction of $-0.18$ from $(w - p - y + l - \hat{\mu})$

Coefficient of $f_t \Delta p_t$ highly significant, but $< 1$

Non-linearity in unemployment is $-0.42U_{t,t}(1 - 6.1U_{t,t})$: negative till unemployment rate exceeds $\approx 15\%$, then positive—only consistent with involuntary unemployment

Step indicators needed to explain higher growth rate of real wages post war (1.9% p.a., versus 0.8% p.a. pre-1945), even though $\Delta(y - l)$ is included and has similar behaviour: spike in 1940 was a permanent location shift

Interactions of variables with step shifts matter as well

Both steps and impulses mainly for wars

(14) encompasses previous models

All mis-specification tests insignificant & constant over 2005–2011

Super exogeneity of $(y - l)_t$, $\Delta p_t$ & $U_{t,t}$ in (14) accepted.
Selected model graphical statistics

Δ \((w - p)\) ^ Δ \((w - p)\) steps

1900 1950 2000

0.0 0.1

a

Δ \((w - p)\) ^ Δ \((w - p)\) steps scaled residuals 'forecast' error

1900 1950 2000

-2 0 2

b

1-step 'forecasts' Δ \((w - p)\)

2005 2010

-0.025 0.000 0.025 0.050

0.2 0.4

c

Residual density N(0,1)

2005 2010

4.75 4.80 4.85

1-step 'forecasts' of \((w - p)_t\)

2005 2010

-0.5 0.0 0.5

0.0 0.2 0.4

d

Non-linear inflation catch-up

\(f_t\)

-0.2 -0.1 0.0 0.1 0.2

0.0 -0.5

e

\(\Delta p\)
Testing exogeneity

- SIS used to test exogeneity of the conditioning variables, extending Hendry and Santos (2010).
- Under null of super exogeneity, parameters in conditional model should be invariant to shifts in marginal models: so indicators in latter should not enter former.

VAR in $\mathbf{w} - \mathbf{p}$, $\mathbf{y} - \mathbf{l}$, $\Delta \mathbf{p}$ and $\mathbf{U}_r$ with SIS at $\alpha = 0.005$; retained indicators in the 3 marginal models tested for significance in (14).

### Super exogeneity tests

<table>
<thead>
<tr>
<th>Variable</th>
<th>null distribution</th>
<th>SIS test on (14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(\mathbf{y} - \mathbf{l})_t$</td>
<td>$F(2, 123)$</td>
<td>0.77</td>
</tr>
<tr>
<td>$\Delta \mathbf{p}_t$</td>
<td>$F(7, 118)$</td>
<td>1.87</td>
</tr>
<tr>
<td>$\mathbf{U}_{r,t}$</td>
<td>$F(14, 111)$</td>
<td>1.37</td>
</tr>
<tr>
<td>Joint</td>
<td>$F(20, 105)$</td>
<td>1.41</td>
</tr>
</tbody>
</table>

No evidence against super exogeneity of $(\mathbf{y} - \mathbf{l})_t$, $\Delta \mathbf{p}_t$ & $\mathbf{U}_{r,t}$ in (14).
(1) Selecting empirical models
(2) Simulating ‘1-cut’ selection
(3) Automatic model extensions: *Autometrics*
(4) Detecting and modelling multiple location shifts
(5) Mis-specification testing and encompassing
(6) Empirical model discovery and theory evaluation
(7) Modelling UK real wages over the last 150 years
(8) **Conclusions**
Little difficulty in eliminating almost all irrelevant variables from the GUM (a small cost of search): HP8 when $N = 145 > T = 139$. **Avoids huge costs from under-specified models.**

When the LDGP retained by *Autometrics* if commenced from it, then a close approximation is generally selected when starting from a GUM which nests that LDGP.

Theory formulations can be embedded in the GUM, to be retained without selection, with no impact on estimator distributions, despite selecting over $N > T$ variables.

**Model selection by *Autometrics* with tight significance levels and bias correction is a successful approach which allows many variables, lags, non-linearities and multiple shifts to be tackled jointly while retaining theory insights.**

All the steps are in place for empirical model discovery jointly with theory evaluation.