Econometrics Spring School 2016
Econometric Modelling

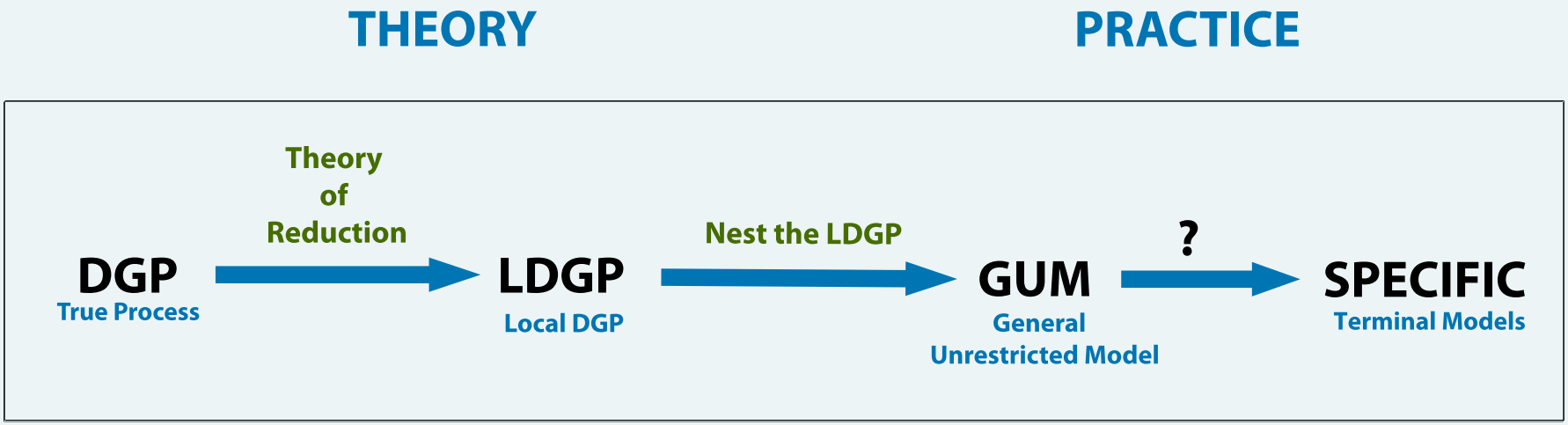
Jurgen A Doornik, David F. Hendry, and Felix Pretis

George-Washington University

March 2016

Lecture 5: Introduction to Automatic Model Selection

GUM to Specific?



How do we move from GUM to specific model?

Selection methods

Many methods for model selection (some frequently used but ineffective in realistic settings).

- Forward selection
- Step-wise regression
- 1-cut elimination
- Backward elimination
- (Best subset selection)
- Information criteria
- Lasso
- (Retina)
- General-to-specific: Gets

Selection methods

Many methods for model selection (some frequently used but ineffective in realistic settings).

- Forward selection
- Step-wise regression
- **1-cut elimination**
- Backward elimination
- (Best subset selection)
- Information criteria
- Lasso
- (Retina)
- **General-to-specific: Gets** (here as *Autometrics*)

Many ways to judge success of selection algorithms

(A) Maximizing the goodness of fit

Traditional criterion for fitting a given model, but does not lead to useful selections

(B) Matching a theory-derived specification

Widely used, and must work well if LDGP $\simeq$ theory, but otherwise need not

(C) Frequency of discovery of the LDGP. Overly demanding–may be nearly impossible even if commenced from LDGP (eg $|\mathrm{t}| < 0.1$)

(D) Improves inference about parameters

Seek small, accurate, uncertainty regions around parameters of interest–but 'oracle principle' invalid

(E) Improved forecasting over other methods
Many contenders: other selections, factors, model averages, robust devices...but forecasting is different

(F) Works for 'realistic' LDGPs
Unclear what those are–but many claimed contenders.

(G) Relative frequency of recovering LDGP starting from GUM as against starting from LDGP
Costs of search additional to commencing from LDGP

(H) Operating characteristics match theory
Nominal null rejection frequency matches actual; retained parameters of interest unbiasedly estimated

(I) Find well-specified undominated model of LDGP
'Internal criterion'–algorithm could not do better

(G), (H) and (I) are main basis: aim to satisfy all three

Two costs of selection: costs of **inference** and **search**
First inevitable if tests have non-zero null and non-unit rejection frequencies under alternative
Applies even if commence from LDGP.
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 the GUM as against the LDGP, plus those for retained irrelevant variables

Also see if *Autometrics* 'outperforms' other automatic methods: **Information Criteria, Step-wise, Lasso, ...**

Probabilities of null rejections in t-testing for N irrelevant regressors at significance level α (critical value c_α):

event	probability	retain
$\mathsf{P}\left(\left\|\mathsf{t}_i\right\| < c_\alpha,\ \forall i = 1, \ldots N\right)$	$(1-\alpha)^N$	0
$\mathsf{P}\left(\left\|\mathsf{t}_i\right\| \geq c_\alpha \mid \left\|\mathsf{t}_j\right\| < c_\alpha,\ \forall j \neq i\right)$	$N\alpha\,(1-\alpha)^{N-1}$	1
$\vdots$	$\vdots$	$\vdots$
$\mathsf{P}\left(\left\|\mathsf{t}_i\right\| < c_\alpha \mid \left\|\mathsf{t}_j\right\| \geq c_\alpha,\ \forall j \neq i\right)$	$N\alpha^{(N-1)}\,(1-\alpha)$	$N-1$
$\mathsf{P}\left(\left\|\mathsf{t}_i\right\| \geq c_\alpha,\ \forall i = 1, \ldots N\right)$	α^N	N

Average number of null variables retained is:

$$k = \sum_{i=0}^{N} i \frac{N!}{i!\,(N-i)!} \alpha^i\,(1-\alpha)^{N-i} = N\alpha. \tag{1}$$

For $N = 40$ when $\alpha = 0.01$ this yields $k = 0.4$

Few spurious variables ever retained, yet 2^N possible models, namely 10^{12}.

Approximate power if coefficient null **only tested once**

t-test powers

ψ	α	$\mathsf{P}(\lvert t\rvert \geq c_\alpha)$	$\mathsf{P}(\lvert t\rvert \geq c_\alpha)^4$
1	0.05	0.16	0.001
2	0.05	0.50	0.063
2	0.01	0.26	0.005
3	0.01	0.64	0.168
4	0.05	0.98	0.902
4	0.01	0.91	0.686
6	0.01	1.00	0.997

50–50 chance of retaining when $\mathsf{E}[t^2] = 4$ for $c_\alpha = 4$
Only 6% chance of keeping **4** such variables

Does repeated testing distort selection?

- (a) Severe illness:
  more tests **increase** probability of **correct diagnosis**.
- (b) Mis-specification tests:
  if r independent tests τ_j conducted under null
  for small significance level η (critical value c_η):

$$\mathsf{P}(|\tau_j| < c_\eta \mid j = 1, \ldots, r) = (1 - \eta)^r \simeq 1 - r\eta.$$

  More tests **increase** probability of **false rejection**.
  Suggests significance level η of 1% or tighter.

Conclude: no generic answer.

Consider a perfectly orthogonal regression model:

$$y_t = \sum_{i=1}^{N} \beta_i z_{i,t} + \epsilon_t \tag{2}$$

$\mathsf{E}[z_{i,t} z_{j,t}] = \lambda_{i,i}$ for $i = j$ & $0\ \forall i \neq j$, $\epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2]$ and $T >> N$.

Order the N sample t^2-statistics testing H_0: $\beta_j = 0$:

$$\mathsf{t}^2_{(N)} \geq \mathsf{t}^2_{(N-1)} \geq \cdots \geq \mathsf{t}^2_{(1)}$$

Cut-off m between included and excluded variables is:

$$\mathsf{t}^2_{(m)} \geq c^2_\alpha > \mathsf{t}^2_{(m-1)}$$

Larger values retained: all others eliminated.
Only one decision needed even for $N \geq 1000$:
'repeated testing' does not occur, and
'goodness of fit' is never considered.

Maintain average false null retention at **one variable** by $\alpha \leq 1/N$, with α declining as $T \to \infty$

Performance of selection algorithms well known
for stationary and ergodic autoregressions:
AIC, *BIC* and *HQ* (Hannan-Quinn) penalize log-likelihood by $f(n,T)$
for n parameters and sample T.

BIC and *HQ* consistent:
DGP$\subseteq$model selected with p $\rightarrow 1$ as $T \rightarrow \infty$ relative to n:
$2n\log(\log(T))/T$ is minimum rate.

**Alternatively, non-centralities ψ diverge, and
significance levels α converge to zero at suitable rate.
Can achieve consistent selection of any finite-sized model.**

Problems with Information Criteria:

- IC do not ensure adequate initial model specification (GETS tests GUM for congruency)
- Selection criteria too loose as N $\rightarrow$ T
- Unclear how to use when $N >> T$

GETS does correct these drawbacks.

Introduction to concepts and practice of model selection

- We shall overview a range of model selection approaches, and their advantages/shortcomings.
- Illustration: very simple biometric application. A range of model selection approaches will be computed by hand.
- Automatic model selection: Autometrics will be described using the simple illustration.

As part of modelling process may need to consider

- model validity
- integration/cointegration
- (weak) exogeneity
- which variables are relevant, which aren't
- functional forms
- non-linearities
- lags
- breaks, outliers
- factors

Much can go wrong, and may need to start all over again

Modern econometric tools can help with many decisions

Selecting relevant variables

Assume decided on

- dependent variable
- possible relevant variables
- type of model (linear in coefficients)
- maximum lag lengths
- sample period

Wish to reduce model to simplify for analysis, forecasting, etc.: remove irrelevant variables.

Any decision introduces probability of mistake

Selecting relevant variables

Assume decided on

- dependent variable
- possible relevant variables
- type of model (linear in coefficients)
- maximum lag lengths
- sample period

Wish to reduce model to simplify for analysis, forecasting, etc.: remove irrelevant variables.

Any decision introduces probability of mistake

Relevant variable can appear irrelevant:

- low statistical significance in the DGP
- not enough variation in sample
- ineffective selection device
- some error prevents us from finding it (wrong model, breaks, outliers, etc.)

We use a simple example to focus on the practical aspects of model selection.

Biological experiment, reported by Finney (1947):
reflex vaso-constriction (tightening of the veins) in the skin of the finger after taking a deep breath:

- response: occurrence of vaso-constriction (Vaso)
- observed 'dose': volume of air inspired (V), rate of inspiration (R)

Inspection of the data indicates that the product of volume and rate governs the response: constant probability along:

$$VR = \text{constant}$$

More generally

$$V^{\alpha} R^{\beta} = \text{constant}$$

which can be linearized as

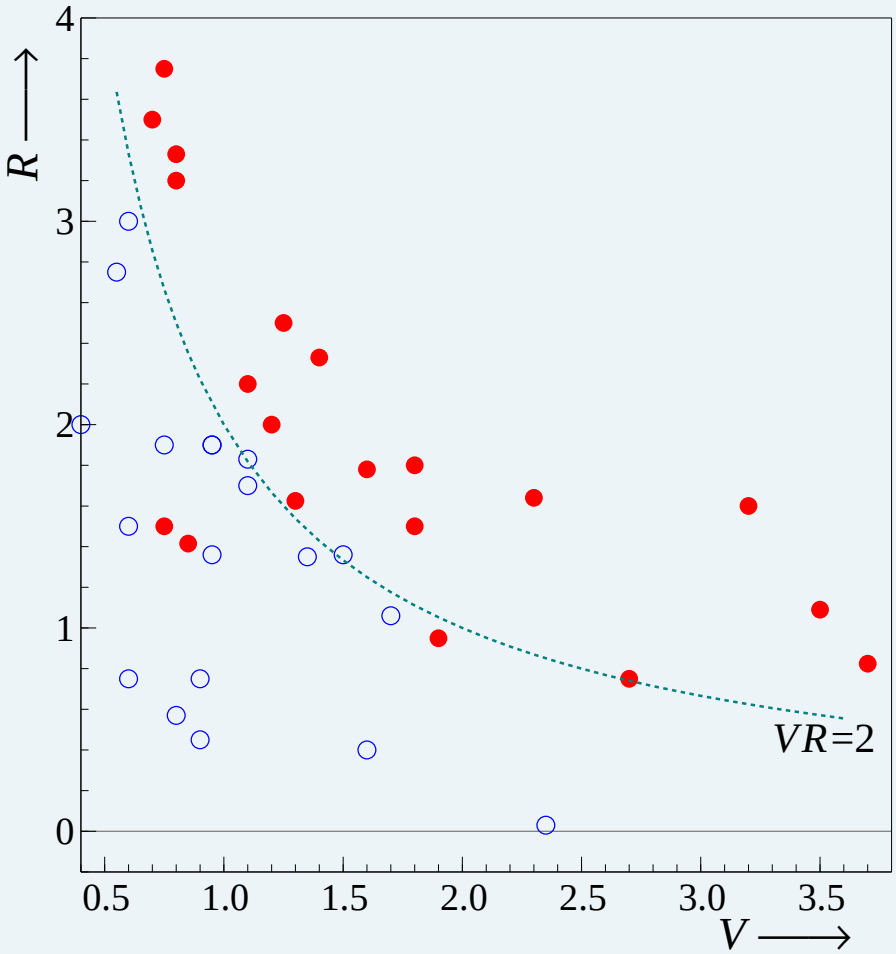
$$\alpha \log V + \beta \log R = \text{constant}$$

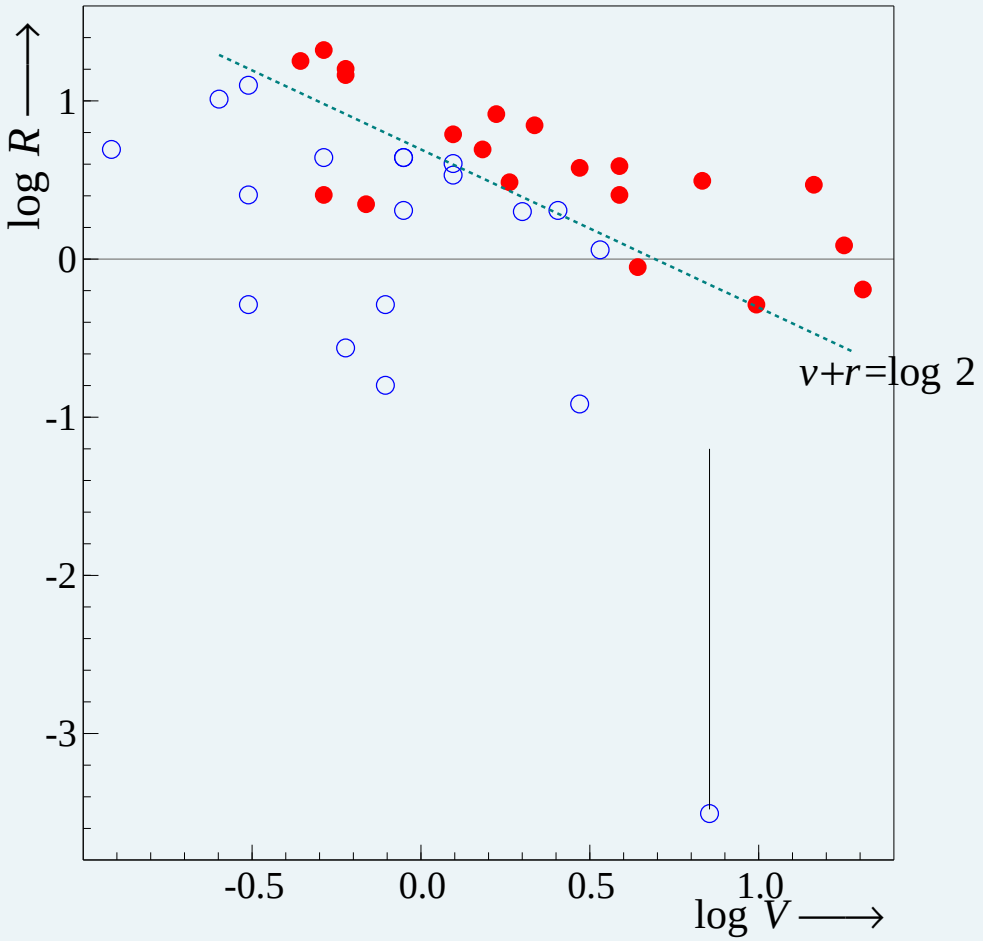
which was the basis for the empirical specification of the probit model (note: computationally challenging then).
Taking a more general starting point:

$$\begin{aligned} \mathcal{X}_{\text{fixed}} &= \{1\} \\ \mathcal{X}_{\text{free}} &= \{v(=\log V), r(=\log R), VR, (VR)^{1/2}\} \end{aligned}$$

using OLS for estimation.

Finney data


R
4
3
2
1
0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
V
VR=2


log R
1
0
-1
-2
-3
-0.5
0.0
0.5
1.0
log V
v+r=log 2

- Load data: `finney.in7`/`finney.bn7`
- Create transformations: $\{v(=\log V), r(=\log R), VR, (VR)^{1/2}\}$.
- Plot data.

Step-wise regression

Forward selection procedure:

- order the regressors, i.e. find most correlated with y,
- add the first regressor
- reorder the remainder
- continue until all significant variables found (at p_a).

Step-wise regression adds:

- after adding a variable, remove the most insignificant variable (at p_a, if any)

Starting point initial set of variables $\mathcal{X}$ (with lags, etc.).

- t-tests (single variable addition, variable removal).

Equivalent approach: select most correlated variable (controlling for variables that are already in the model.

Can use computationally efficient implementations, but not so important these days.

Manual implementation of step-wise regression

- Step-wise regression,

$$\begin{aligned}\mathcal{Y} &= \{\text{vaso}\},\\ \mathcal{X}_{\text{fixed}} &= \{1\},\\ \mathcal{X}_{\text{free}} &= \{r = \text{Lrate}, v = \text{Lvolume}, VR = \text{rate} \times \text{volume}, (VR)^{1/2}\}\end{aligned}$$

  - Compute t-statistics for model with one variable.
  - Add most significant variable and compute t-statistics for model with each additional variable.
  - Continue until no significant variables.
- Significance level: $\alpha = 5\%$.

Choose $p_a = 0.05$, critical value ≈ 2:

Model 1: $\widehat{\text{Vaso}}_i = 0.513$

t value when adding one variable in turn:

r	v	VR	$(VR)^{1/2}$
2.08*	2.71**	**5.83****	5.67**

Model 2: $\widehat{\text{Vaso}}_i = -0.133 + 0.322\,VR_i$

t value when adding one more variable

r	v	VR	$(VR)^{1/2}$
-0.200	0.324	—	0.264

Final model: $\widehat{\text{Vaso}}_i = -0.133 + 0.322\,VR_i$

7 'tests' lead to final results

Advantages:

- Fast and simple
- can handle more variables than observations ($k > T$)

Problems:

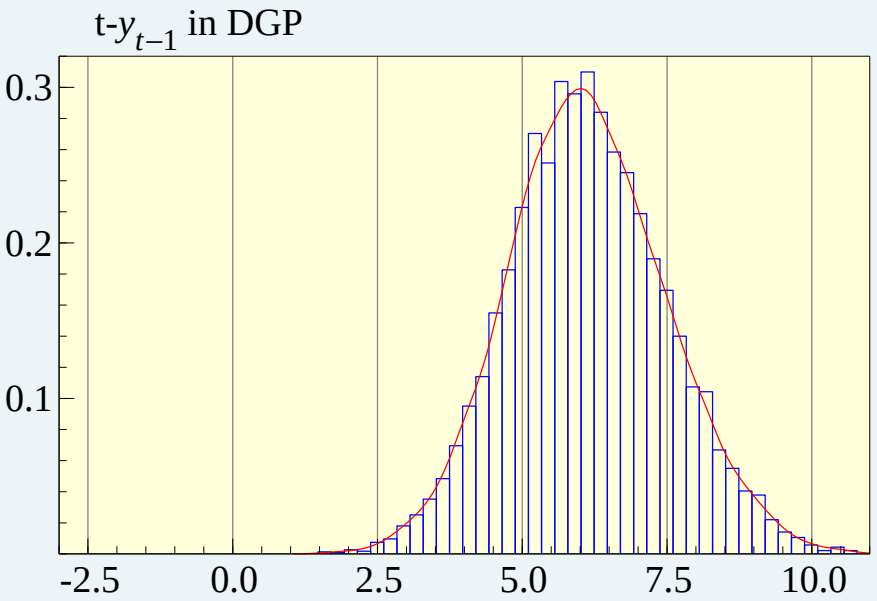
- the ordering can prevent relevant regressors from entering, because they never appear towards the front (until the model gets large).
  Example: x_1 and x_2 are negatively correlated and feature in the model together. However, individually they will not rank high on the ordering method.
- ordering criteria may be biased at an early stage by the omission of relevant factors. what is the distribution of the test?
- single path creates path dependence: adding insignificant regressors may change the path and thus the final model

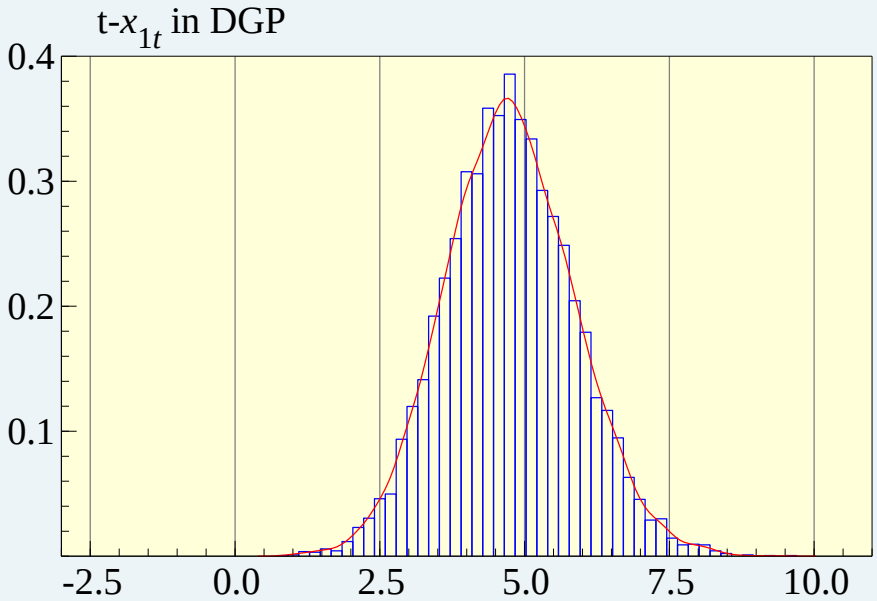
DGP –

$$
\begin{aligned}
y_t &= 0.5y_{t-1} + 0.8x_{1,t} + 0.8x_{2,t} + u_t, \quad u_t \sim \mathsf{IN}[0, 1], \\
\mathbf{x}_t &= \mathbf{v}_t, \qquad\qquad\qquad\qquad\quad \mathbf{v}_t \sim \mathsf{IN}\left[\mathbf{0}, \begin{matrix} 1 & -0.8 \\ -0.8 & 1 \end{matrix}\right],
\end{aligned}
$$

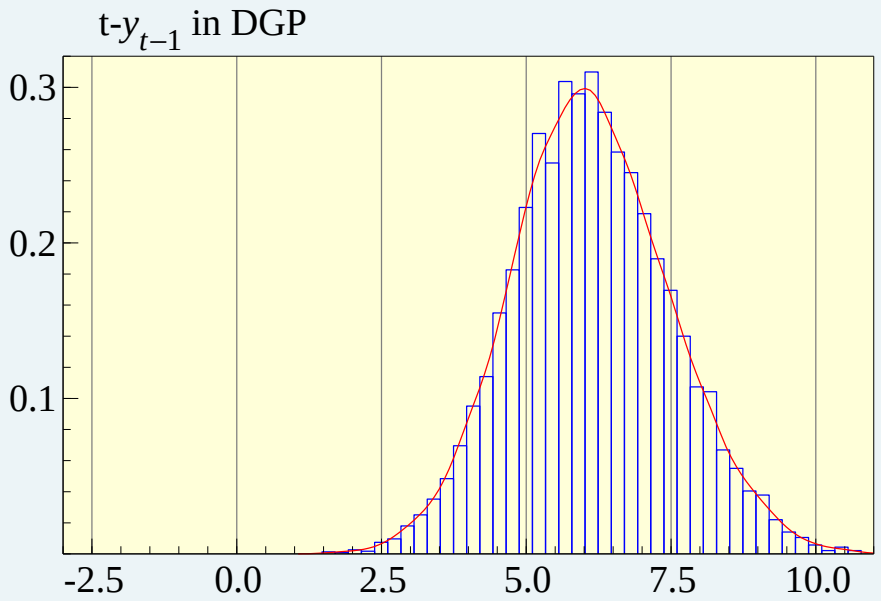
$$
\begin{aligned}
\mathcal{X}_{\mathsf{fixed}} &= \{1\} \\
\mathcal{X}_{\mathsf{free}} &= \{y_{t-1}, x_{1t}, x_{2t}\}
\end{aligned}
$$

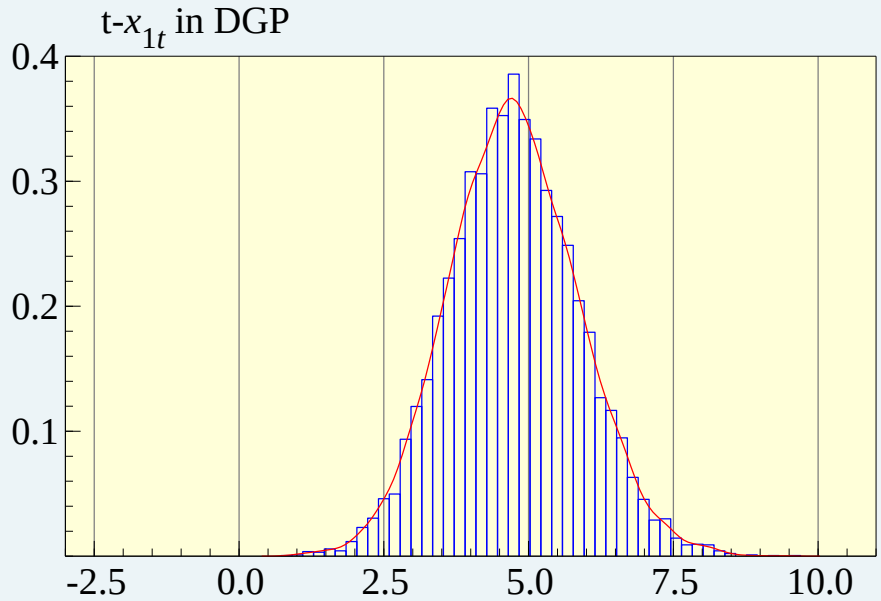
Step-wise t-values vs DGP

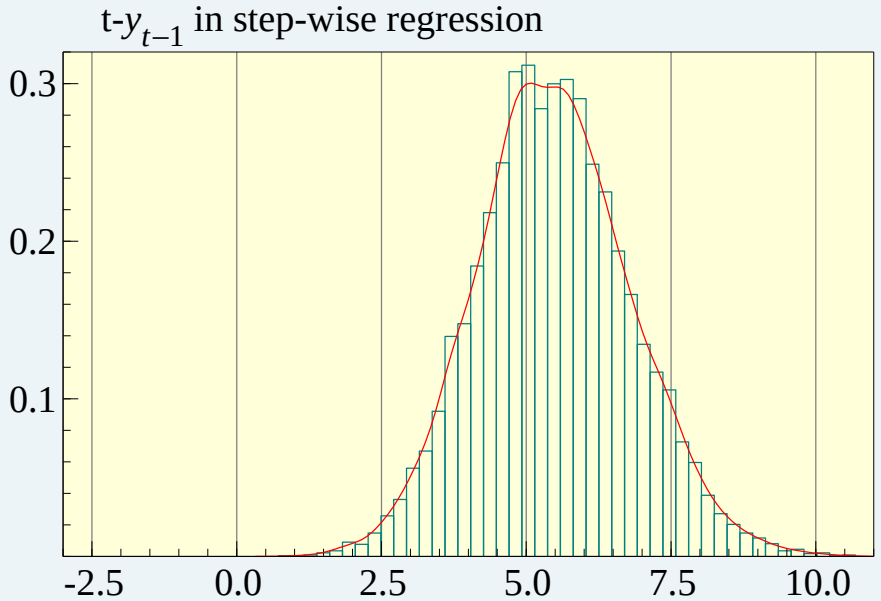


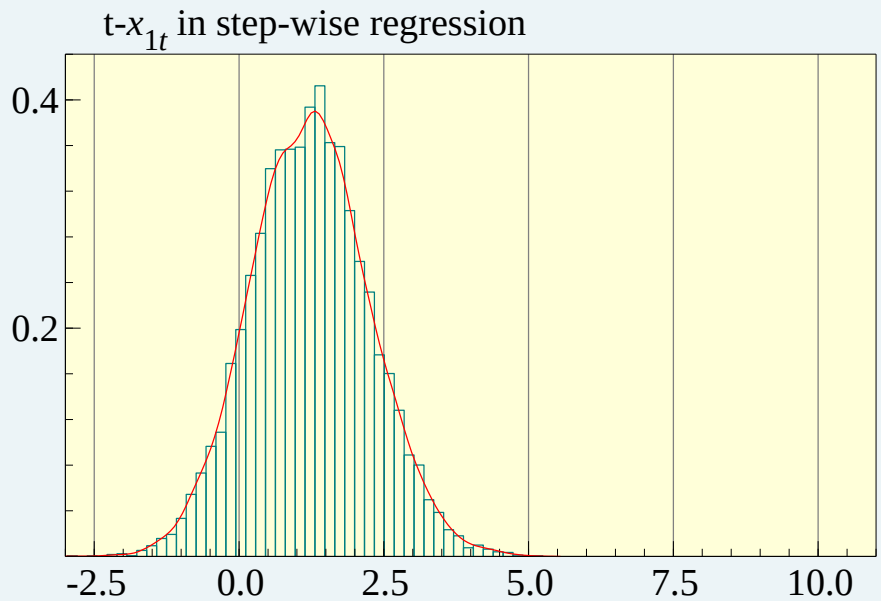


Step-wise t-values vs DGP









1-cut elimination

Delete all variables with t-value such that significance is below p_a.

Order variables by significance, then do a single cut at p_a.

Obviously a bad idea, unless regressors are uncorrelated.

Manual implementation of 1-cut at $p_a = 5\%$.

- Estimate model with all regressors included.
- Order regressors based on their t-statistics.
- Eliminate all regressors with t-statistics less than corresponding critical value.

Choose $p_a = 0.05$, critical value ≈ 2:

Model 1: $\widehat{\text{Vaso}}_i = -2.27 - 0.61 r_i - 0.58 v_i - 0.33\, VR_i + 2.75 (VR)_i^{1/2}$

t value of each free variable

r	v	VR	$(VR)^{1/2}$
-0.905	-0.790	-0.511	0.986

Final model: $\widehat{\text{Vaso}}_i = 0.51$

Backward elimination

Can run stepwise procedure backward:

- Start with all variables in the model,
- Delete the most insignificant, one at a time
- Stop when none left

Ingredients for stepwise regression:

- t-tests (single variable removal).

Run at backward elimination $p_a = 5\%$.

- Compute t-statistics for model with all variables.
- Delete least significant variable and re-estimate model.
- Continue until no insignificant variables.

Backward elimination: example

Choose $p_a = 0.05$, critical value ≈ 2:

Model 1: $\widehat{\text{Vaso}}_i = -2.27 - 0.61 r_i - 0.58 v_i - 0.33 VR_i + 2.75 (VR)_i^{1/2}$

	r	v	VR	$(VR)^{1/2}$
t values	-0.905	-0.790	**-0.511**	0.986

Model 2: $\widehat{\text{Vaso}}_i = -1.19 - 0.31 r_i - 0.27 v_i + 1.36 (VR)_i^{1/2}$

	r	v	VR	$(VR)^{1/2}$
t values	-0.970	**-0.649**	—	2.11

Model 3: $\widehat{\text{Vaso}}_i = -0.753936 - 0.11 r_i + 0.96 (VR)_i^{1/2}$

	r	v	VR	$(VR)^{1/2}$
t values	**-1.17**	—	—	5.15

Final model: $\widehat{\text{Vaso}}_i = -0.60 + 0.82 (VR)_i^{1/2}$

4 'tests' lead to final results

Backward elimination: pros/cons

Can run stepwise procedure backward:

- All variables are in at first, so will maintain pairs when necessary,
- the starting point may be statistically better behaved.

But:

- still only one path, therefore:
- path dependence: adding insignificant regressors will change the path and may change the final model

Performance of selection by IC well known for stationary and ergodic autoregressions:

- *AIC*, *SC* and *HQ* penalize log-likelihood by $f(k,T)$ for k parameters and sample T.
- *SC* and *HQ* consistent:
  DGP$\subseteq$model selected with prob $\to 1$ as $T \to \infty$ relative to k:
  $2kT^{-1}\log(\log(T))/T$ is minimum rate.
- Need to estimate all 2^k models to properly minimize information criterion.

$$\begin{aligned}
\text{SC} &= \left(-2\widehat{\ell} + k\log T\right)T^{-1}, \\
\text{HQ} &= \left(-2\widehat{\ell} + 2k\log\log T\right)T^{-1}, \\
\text{AIC} &= \left(-2\widehat{\ell} + 2k\right)T^{-1},
\end{aligned}$$

Use progress to select by IC from all models estimated sofar.

Lasso is OLS with a constraint on the absolute sum of the parameters.

$$y_t = \beta_0 + \sum_{j=1}^{k} x_{jt}\beta_j + \varepsilon_t, \quad t = 1, ..., T.$$

Rewrite in deviation from mean:

$$y_t^* = \sum_{j=1}^{k} x_{jt}^*\beta_j + \varepsilon_t,$$

with residuals

$$r_t(\beta_1, ..., \beta_k) = y_t^* - \sum_{j=1}^{k} x_{jt}^*\beta_j.$$

Lasso solves:

$$(\widehat{\beta}_1^{L(C)}, ..., \widehat{\beta}_k^{L(C)}) = \text{argmin}_\beta \sum_{t=1}^{T} r_t^2(\beta_1, ..., \beta_k) \text{ s.t. } \sum_{j=1}^{k} |\beta_j| \leq C.$$

- If no binding constraint: equal to OLS,
- otherwise parameters shrunk towards 0.
- initially solve as QP problem, but Efron, Hastie, Johnstone and Tibshirani (2004) provide elegant and fast algorithm.
- Parsimony: only subset of βs nonzero when constraint is binding.
- 'less greedy' version of step-wise regression

Lasso model selection:

- Run Lasso solution forward until OLS (or perfect fit)
- Choose the Lasso solution with the smallest $C_p(\boldsymbol{\beta}^L)$ or $\mathrm{SC}(\boldsymbol{\beta}^L)$, denote this $\widehat{\beta}_1^{L(*)}, ..., \widehat{\beta}_k^{L(*)}$.
- Re-estimate the model with OLS, keeping only the variables that have $\widehat{\beta}_j^{L(*)} \neq 0$.

Step	size	non-zero coefficients	SC	C_p
1	1	VR	-1.7956	4.9141
2	2	VR, $(VR)^{1/2}$	**-1.8370**	**1.6087**
3	3	VR, $(VR)^{1/2}$, v	-1.7640	2.8523
4	3	VR, $(VR)^{1/2}$, r	-1.7684	2.6950
5	3	$(VR)^{1/2}$, r, v	-1.7783	2.3432
6	4	$(VR)^{1/2}$, r, v, VR	-1.6941	4.0000

Final model: $\widehat{\text{Vaso}}_i = -0.24 + 0.18(VR)_i^{1/2} + 0.26 VR_i$

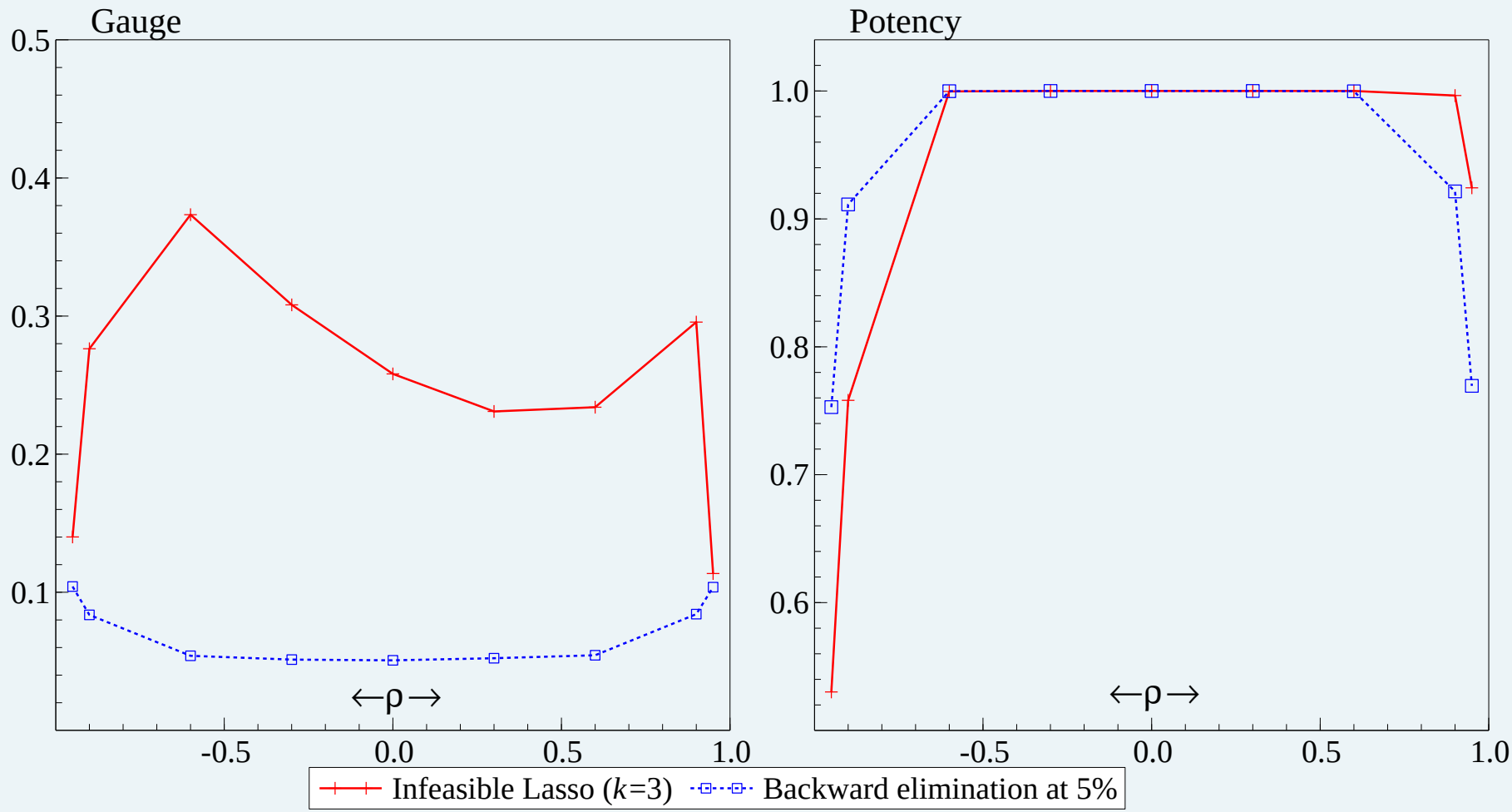
- Single path,
- Difficult to know when to stop.

DGP –

$$
\begin{aligned}
y_t &= 0.5y_{t-1} + 0.8x_{1,t} + 0.8x_{2,t} + u_t, \quad u_t \sim \mathsf{IN}[0,1], \\
\mathbf{x}_t &= \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathsf{IN}\left[\mathbf{0}, \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}\right],
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{X}_{\text{fixed}} &= \{1\} \\
\mathcal{X}_{\text{free}} &= \{y_{t-1}, x_{1t}, x_{2t}, x_{3t}, x_{4t}\}
\end{aligned}
$$

Lasso vs backward elimination



gauge: fraction of irrelevant variables (x_{3t}, x_{4t}) in the final model

potency: fraction of relevant variables $(y_{t-1}, x_{1t}, x_{2t})$ in final model.

- Define a starting model: general unrestricted model (GUM)
  - Designed to be congruent (diagnostic testing) and relevant,
  - Tests of reductions with approximately correct distribution,
  - Reduction can maintain congruence (or lack thereof),
  - Reduction up to a predefined significance level (backtesting w.r.t. GUM: acceptable information loss).

- Define a starting model: general unrestricted model (GUM)
  - Designed to be congruent (diagnostic testing) and relevant,
  - Tests of reductions with approximately correct distribution,
  - Reduction can maintain congruence (or lack thereof),
  - Reduction up to a predefined significance level (backtesting w.r.t. GUM: acceptable information loss).

Model selection is an iterative search procedure, need to follow several paths:

- multiple path search, or
- tree search.

- t-tests (single variable removal).
- F-tests (tests of variables removed from the GUM, encompassing aka backtesting).
- F-tests (pruning to faster search).
- diagnostic tests
  - ARCH (Engle 1982)
  - Serial correlation (Godfrey 1978, Harvey 1981)
  - Heteroscedasticity (White 1980)
  - Normality (Jarque and Bera 1980; Doornik and Hansen 1994, 2008)
  - Chow (Chow 1960 in-sample stability test)
- information criterion (tiebreaker)
- stability tests (out of sample, optionally)

Stage 0: formulate and estimate GUM (general unrestricted model)
Stage 1: diagnostic testing of GUM

```
Normality test: Chi^2(2) = 0.29810 [0.8615]
Hetero test:    F(7,31)  =  4.6698 [0.0012]**
Hetero-X test:  F(13,25) =  2.8003 [0.0131]*
RESET test:     F(1,33)  =  2.5130 [0.1225]
```

Problem: heteroscedasticity
Not surprising: should have used logit/probit.

Normally: need to reconsider GUM. In this case: ignore.

Current example

There are four insignificant variables in the GUM
Putting most insignificant first: VR, v, r, $(VR)^{1/2}$.

Automated Gets

Current example

There are four insignificant variables in the GUM
Putting most insignificant first: VR, v, r, $(VR)^{1/2}$.

Automated Gets

- Delete one insignificant variable in the GUM

Current example

There are four insignificant variables in the GUM
Putting most insignificant first: VR, v, r, $(VR)^{1/2}$.

Automated Gets

- Delete one insignificant variable in the GUM
- then do backward elimination on the reduced model,

Current example

There are four insignificant variables in the GUM
Putting most insignificant first: VR, v, r, $(VR)^{1/2}$.

Automated Gets

- Delete one insignificant variable in the GUM
- then do backward elimination on the reduced model,
- augmented with encompassing (backtesting),

Current example

There are four insignificant variables in the GUM
Putting most insignificant first: VR, v, r, $(VR)^{1/2}$.

Automated Gets

- Delete one insignificant variable in the GUM
- then do backward elimination on the reduced model,
- augmented with encompassing (backtesting),
- and with diagnostic tracking.

Current example

There are four insignificant variables in the GUM
Putting most insignificant first: VR, v, r, $(VR)^{1/2}$.

Automated Gets

- Delete one insignificant variable in the GUM
- then do backward elimination on the reduced model,
- augmented with encompassing (backtesting),
- and with diagnostic tracking.
- Now return to GUM, delete next insignificant variables and repeat process.

This defines four backward elimination paths:

path	*insignificant variables*	F$(3,34)$-*test*	*final*
VR	VR, v, r	0.66449 [0.5797]	$(VR)^{1/2}$
v	v, VR, r	0.66449 [0.5797]	$(VR)^{1/2}$
r	r, $(VR)^{1/2}$, v	0.36828 [0.7763]	VR
$(VR)^{1/2}$	$(VR)^{1/2}$, r, v	0.36828 [0.7763]	VR

Final model 1: $\widehat{\text{Vaso}}_i = -0.60 + 0.82(VR)_i^{1/2}$

Final model 2: $\widehat{\text{Vaso}}_i = -0.13 + 0.32 VR_i$

At first sight 16 t-tests and 4 F tests.

- t-tests are mainly used to order variables for F tests
- some F test are repeated

More like two 1-cut eliminations, one for each final model.

p_a controls information loss relative to GUM:

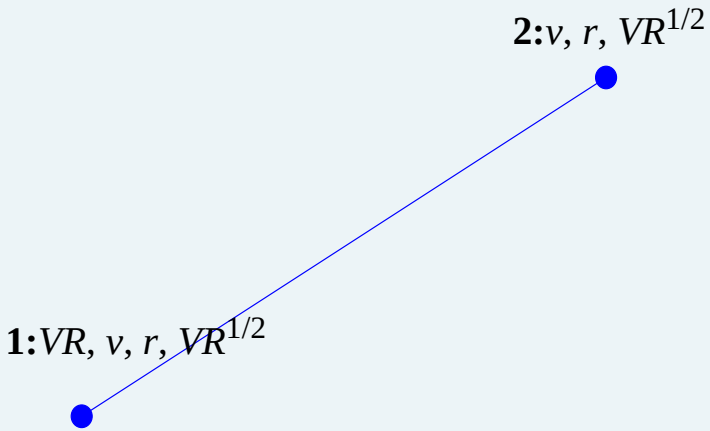
- no attempt to maximize fit
- instead: find parsimonious model up to tolerated loss of fit.
- diagnostic tracking (here Hetero test ignored)
  - may need to backtrack from terminal candidate
- the parsimonious encompassing F test may fail
  - may need to backtrack from terminal candidate

Two reasons why a final model may have insignificant variables.

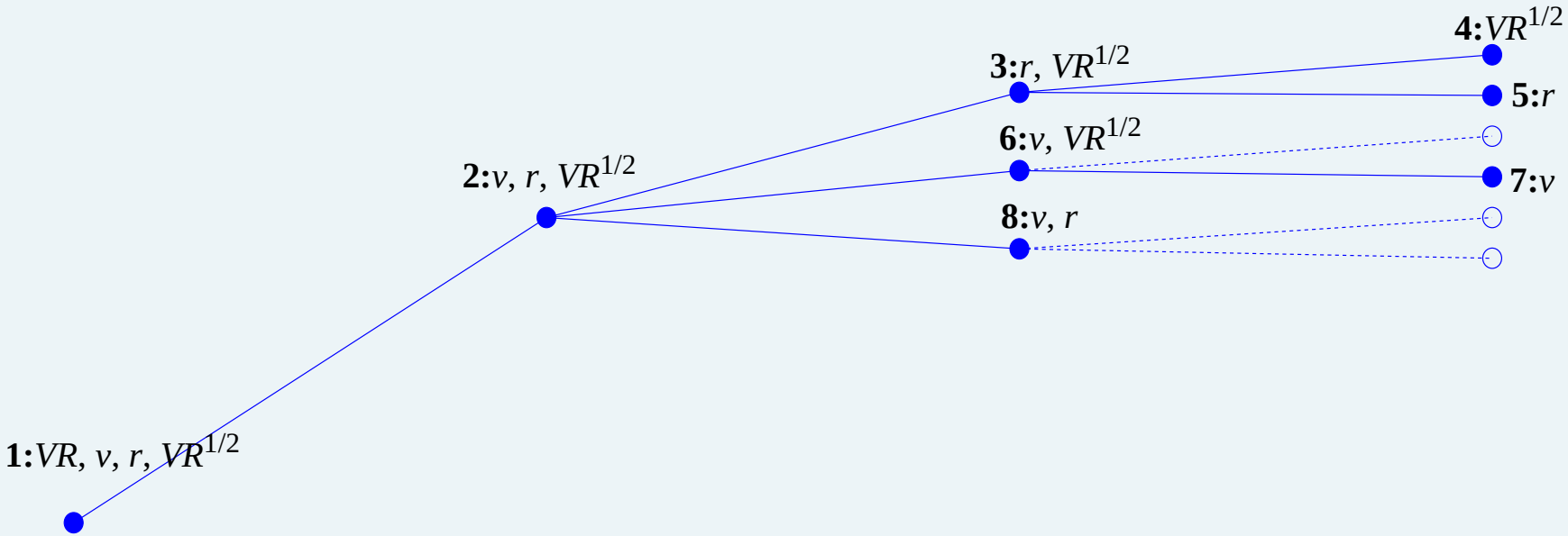
More efficient to use a tree representation of the model space:

- Multiple path search can miss reduction paths
- Avoids re-estimating the same model twice (or more)
- Only need to remember one reduction path from the GUM along which it is possible to backtrack
- Can implement short-cuts along a path to speed up search

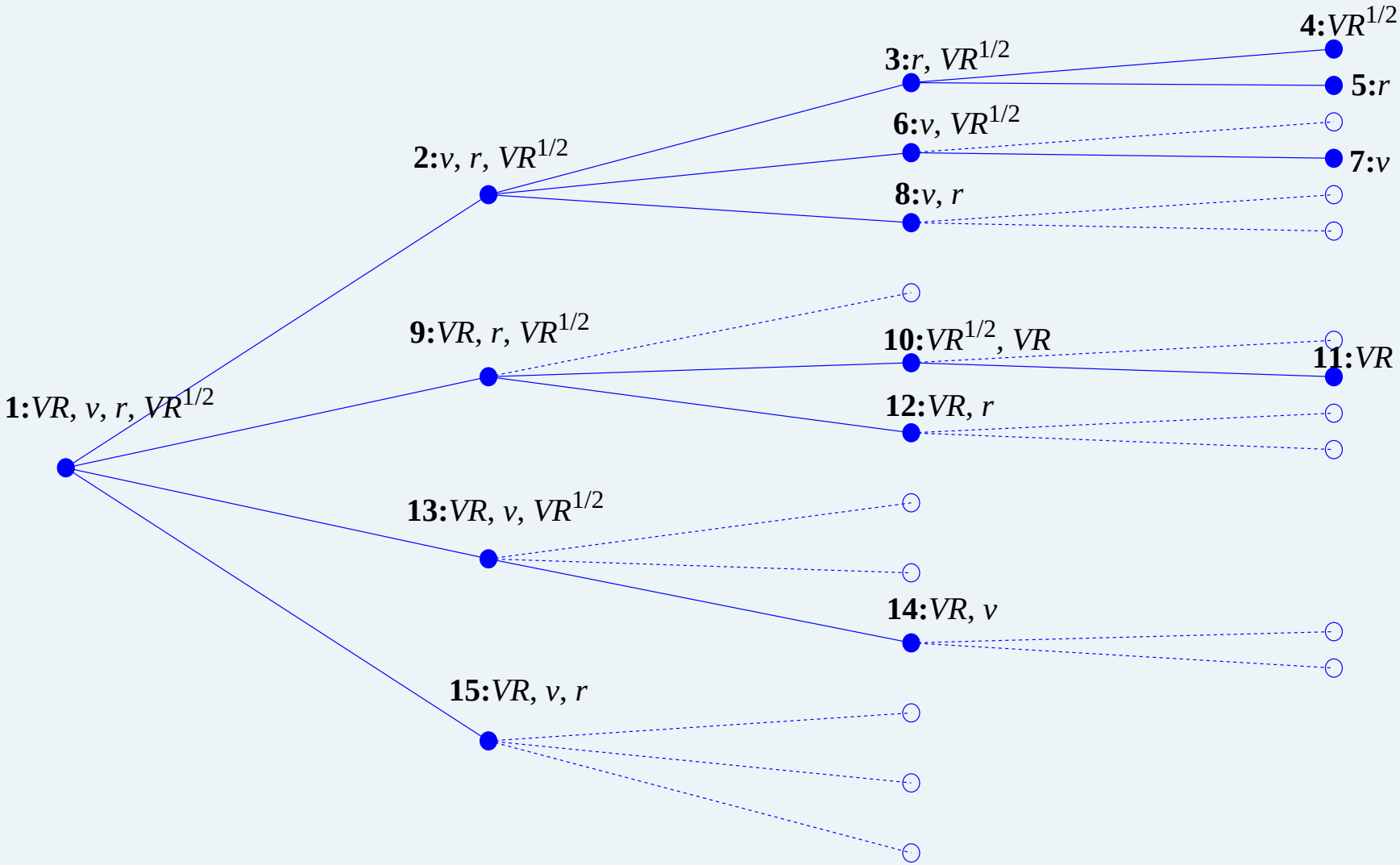
Gets: tree search



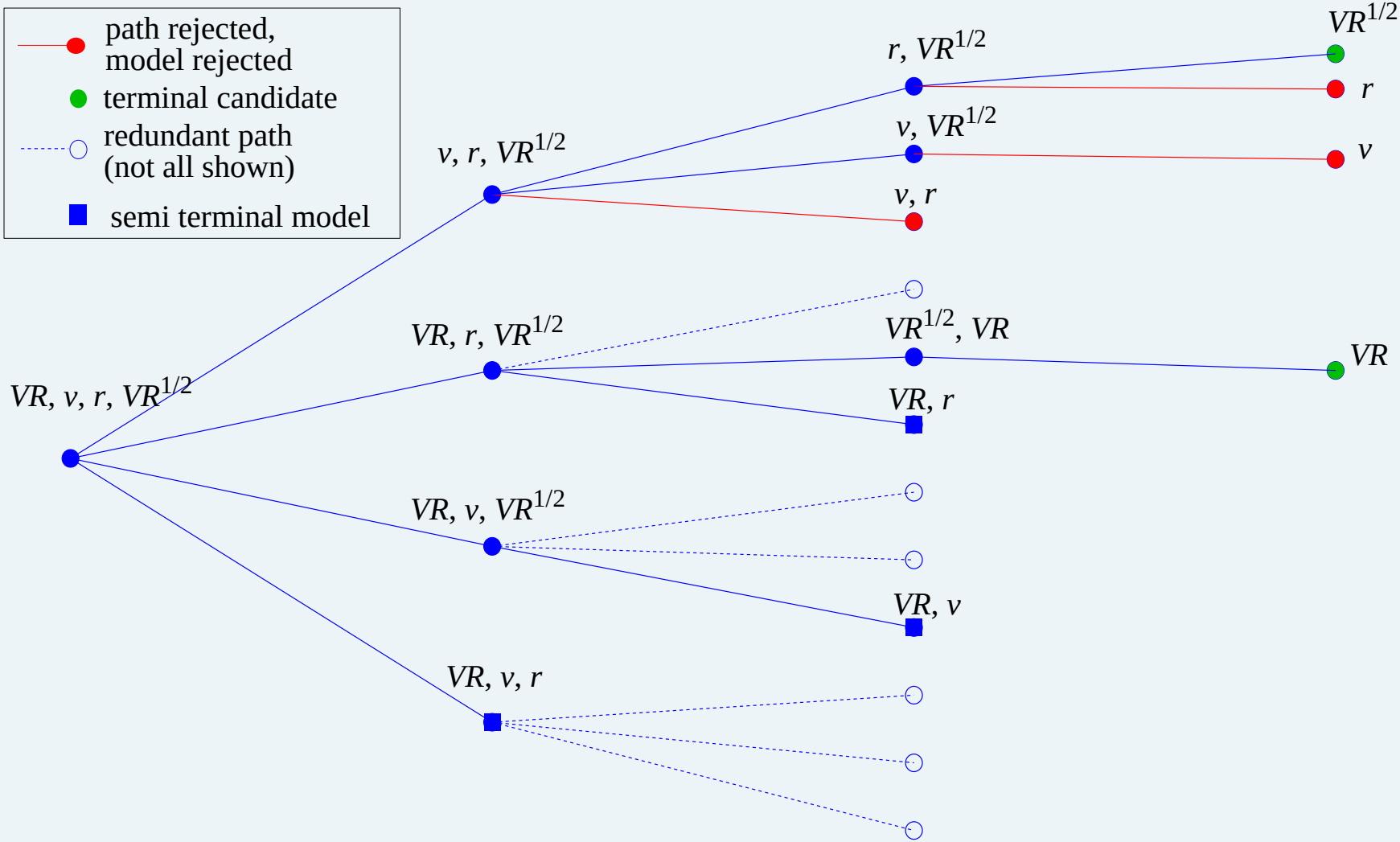
Gets: tree search



Gets: tree search



Gets: tree search



- v,r not a valid reduction of the GUM (Finney's model),
- Once VR and $(VR)^{1/2}$ have been found as terminal models: all further models with these in are redundant (because they nest the terminal, which was already established as a valid reduction).
- Estimated 8 models (instead of 16 for multiple path search)

May have economic or other reasons to prefer one terminal model over another.

Otherwise could use information criterion to choose:

p-values in Final GUM and terminal model(s)

	Final GUM	terminal 1	terminal 2
VR	0.3281	.	0.00000079
$(VR)^{1/2}$	0.7934	0.00000000	.
k	2	1	1
parameters	3	2	2
loglik	-15.563	-16.088	**-15.600**
AIC	0.95194	0.92759	**0.90259**
HQ	0.99785	0.95820	**0.93320**
SC	1.0799	1.0129	**0.98790**

- Model selection is an iterative search procedure
  - manual search can follow a few paths: slow and tedious,
  - computer automated search can follow all paths,
    Well, not all. There are 2^k models, so need a strategy.
    $k = 100$ at 10^9/sec: $10^6\times$ age of universe.

- Model selection is an iterative search procedure
  - manual search can follow a few paths: slow and tedious,
  - computer automated search can follow all paths,
    Well, not all. There are 2^k models, so need a strategy.
    $k = 100$ at 10^9/sec: $10^6\times$ age of universe.

- General-to-specific model selection (Gets, 'Hendry' or 'LSE' methodology) largely driven by David Hendry (DHSY, PcGive, Alchemy, Dynamic Econometrics, ...)
  Lively debate.
- Automated Gets initiated by Hoover and Perez (1999), Hendry and Krolzig (2005) (PcGets: 2nd generation, theoretical properties, bias correction).
  Study model selection through simulation – improves debate.
- Autometrics (Doornik (2009), 3rd generation) improves on PcGets, extended beyond regression models.

Features of GETS algorithms

- Hoover-Perez (1999):
  1. General unrestricted model
  2. Multiple path search
  3. Encompassing test
  4. Diagnostic testing
  5. Tiebreaker
- Hendry and Krolzig (1999), PcGets (2001):
  1. Add presearch
  2. Extend multiple-path search
  3. Add iteration
  4. No out-of-sample testing (Lunch and Vital-Ahuja, 1998)
  5. Change treatment for Invalid GUM
- PcGive (2007) Doornik (2008), Autometrics (2009):
  1. Reduce role of presearch
  2. Change search path algorithm: tree search
  3. Extend scope: separation of model and algorithm
  4. Increase efficiency

Autometrics implements underlying principle of general-to-specific model selection (*'Hendry methodology'*).

Autometrics

- likelihood-based: separation of model and selection
- Only using lag presearch (by default)
- searches the whole model space:
  - tree search ensures that no model is estimated twice
  - irrelevant paths can be cut-off efficiently
  - F-tests are used to speed-up search
- implements backtracking on diagnostics: only test from terminal candidates, then backtrack if necessary
- backtesting w.r.t. GUM 0 (the initial GUM after presearches) removes need for encompassing of candidate models
- relevant terminal candidates remembered in iterated search
- implements block search for $k \geq T$ ($k \geq \alpha T$)

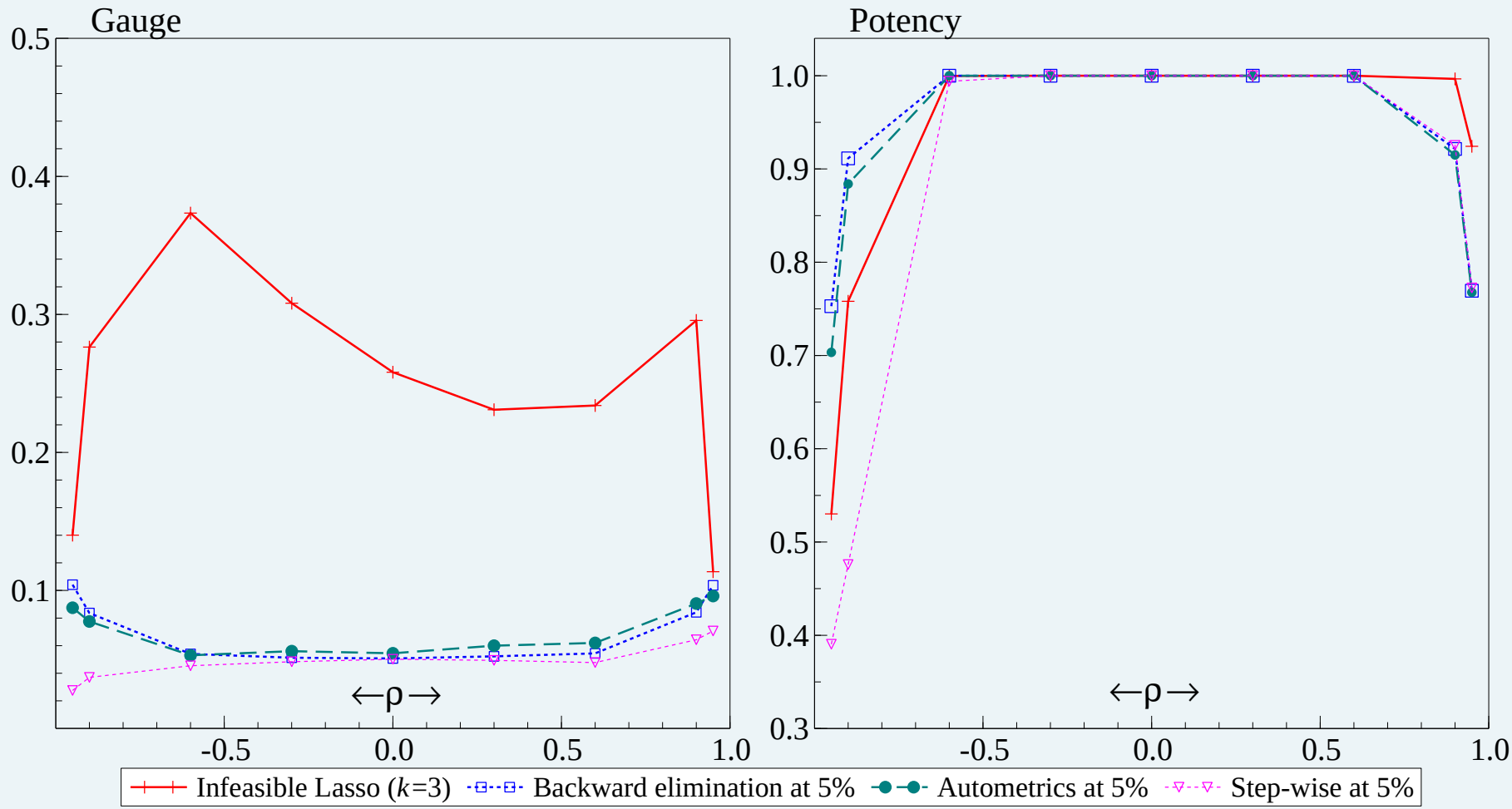
DGP –

$$
\begin{aligned}
y_t = & 0.5y_{t-1} + 0.8x_{1,t} + 0.8x_{2,t} + u_t, \quad u_t \sim \mathsf{IN}[0,1], \\
\mathbf{x}_t = & \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathsf{IN}\left[\mathbf{0}, \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}\right],
\end{aligned}
$$

Model –

$$
\begin{aligned}
\mathcal{X}_{\text{fixed}} &= \{1\} \\
\mathcal{X}_{\text{free}} &= \{y_{t-1}, x_{1t}, x_{2t}, x_{3t}, x_{4t}\}
\end{aligned}
$$

Four selection methods



gauge: fraction of irrelevant variables (x_{3t}, x_{4t}) in the final model

potency: fraction of relevant variables $(y_{t-1}, x_{1t}, x_{2t})$ in the final model.

Many methods of model selection available.

Only rarely justification for specific-to-general (stepwise):

- lack of identification
- only works when there is no path dependence (approximate independence or single nested sequence)

General-to-specific model selection shown to have better properties across wide range of states of nature.

Automation of model selection:

- better tools lead to better empirical models
- Making model selection more objective: different researchers more likely to obtain same results.

Doornik, J. A. (2009).
Autometrics.
In J. L. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics: Festschrift in Honour of David F. Hendry*. Oxford: Oxford University Press.

Hendry, D. F. and H.-M. Krolzig (2005).
The properties of automatic Gets modelling.
Economic Journal 115, C32–C61.

Hoover, K. D. and S. J. Perez (1999).
Data mining reconsidered: Encompassing and the general-to-specific approach to specification search.
Econometrics Journal 2, 167–191.