Forecasting by factors, by variables, by both, or neither?

Jennifer L. Castle[*], Michael P. Clements[†]
and David F. Hendry[*]

[*] **Institute for New Economic Thinking**

at the Oxford Martin School, University of Oxford

[†] **Economics Department, Warwick University**

George Washington University, September 2011

Overview

Is it better to forecast by factors, by variables, by both, or by neither?

Select forecasting models from candidate set including both regressors and their principal components jointly.

Requires selection method to handle perfect collinearity & more regressors than observations.

Autometrics allows that, and tackles breaks and outliers by also including an indicator variable for every observation.

Apply to forecasting US GDP and inflation over 1-, 4- and 8-step horizons:
factor models more useful for 1-step ahead;
variables tend to be better as the horizon increases;
difficult to beat autoregressions.

Background

Three venerable traditions in economic forecasting

[1] Economic-theory derived empirical econometric models:
Klein (1950), Wallis (1989), Smets and Wouters (2003)

[2] Factor approaches combining many sources of information:
Persons (1924), Zarnowitz and Boschan (1977), Stock and Watson (1989, 1999, 2009), Forni, Hallin, Lippi and Reichlin (2000)

[3] Mechanistic approaches:
Holt (1957), Winters (1960), Box and Jenkins (1970).

Few direct comparisons of first two, none including both.
Most comparisons to third,

Route map

(1) **Setting the scene**

(2) Relating variables $\{\mathbf{z}_t\}$ to factors $\{\mathbf{f}_t\}$

(3) Comparing variable-based and factor-based models

(4) Taxonomy of forecast errors in factor models

(5) Forecast failure in equilibrium-correction models

(6) Model selection with both factors and variables

(7) Forecasting US GDP and inflation

Conclusions

Setting the scene

Eight interacting issues affect analyses of forecasting: show answer to our question is context specific:

(i) include both variables & factors in forecasting models;

(ii) role of in-sample model selection in that setting;

(iii) whether breaks are anticipated or not;

(iv) role of more versus less information in forecasting;

(v) type of forecasting model (equilibrium-correction mechanisms–EqCMs);

(vi) data measurement errors near the forecast origin;

(vii) how to evaluate 'success or failure' of forecasts;

(viii) nature of DGP itself.

Consider these eight below.

(i) Pooling information

Factor models pool many predictors:
or could pool forecasts over many models.

Automatic model selection allows regression to pool information.

Form principal components from n 'external' variables, $\mathbf{z}_t$, where $n > T$ observations.
Let $\mathbf{f}_t = \mathbf{B}\mathbf{z}_t$, where $\mathbf{B}$ is matrix of eigenvectors, then $\mathbf{f}_t, \ldots \mathbf{f}_{t-s}, \mathbf{z}_t, \ldots \mathbf{z}_{t-s}$ are possible candidate variables.

Could also include other models' forecasts as can handle $n > T$ and perfect collinearity.

(ii) Model selection

Multi-path selection with expanding and contracting block searches can handle perfect collinearity and $n > T$, so not insurmountable to include all $2sn$.
Autometrics within *PcGive* (see Doornik, 2009) formulates a general unrestricted model (GUM), checking congruence once $N \ll T$ and no perfect collinearity.

Location shifts and outliers removed by impulse-indicator saturation (IIS: see Johansen and Nielsen, 2009, and Castle, Doornik and Hendry, 2011a).
$\{1_{\{j=t\}}, t = 1, \ldots, T\}$ is complete set of T impulse indicators
initial candidates: $\mathbf{f}_t, \ldots \mathbf{f}_{t-s}$, $\mathbf{z}_t, \ldots \mathbf{z}_{t-s}$, $\{1_{\{j=t\}}, t = 1, \ldots, T\}$
so $N > T$ always occurs when IIS is used.

(iii) Unanticipated location shifts

Ex ante forecasting fundamentally different from *ex post* modeling when unanticipated location shifts occur.
Can always model shifts after the event,
but forecast failure when not anticipated–
location shifts most pernicious for forecasting.

Castle, Fawcett and Hendry (2011b) investigate how to forecast breaks, and if fail, how to forecast during breaks: pessimistic due to limited information at location shift occurrence.

Allow for unanticipated location shifts here.

(iv) Role of information in forecasting

Taxonomy of forecast errors in Clements and Hendry (2005) shows incomplete information *per se* not explanation of forecast failure
–unless that information would forecast breaks.

More information cannot lower predictability,
omitting explanatory variables leads to inferior forecasts,
but large amounts of data will not fix key problem of location shifts, unless can forecast them in advance.

Makridakis and Hibon (2000) conclude parsimonious models best in forecasting competitions,
but parsimonious models often robust to location shifts:
jury still out on benefits of more versus less information when forecasting.

(v) Equilibrium-correcting behavior

Factor models are often EqCMs:
suffer from non-robustness of that class to location shifts.

But principles of robust-model formulation from Hendry (2006) apply:
any equilibrium-correction system could be differenced prior to forecasting.

How a model is used in forecasting matters.

(vi) Measurement errors

Many ‘solutions’ to forecast failure exacerbate adverse effects of measurement errors near the forecast origin: e.g., differencing doubles their impact.

Averaging mitigates effects of random measurement errors, so factors might reduce such errors.

Models that account for data revisions also a solution.

Frequency of macroeconomic data affects its accuracy, as does nowcasting & ‘real time’ (versus *ex post*) forecasting

Data measurement errors seem larger in latest data, and during turbulent periods:
might favour factor models over approaches that ignore data revisions.

(vii) Forecast evaluation

Vast literature on how to evaluate *forecasts*,
and using forecasts to evaluate *models*,
forecasting *methods*,
and economic theory.

Here just report descriptive statistics of forecast performance.

(viii) Nature of the DGP

Nature of the DGP greatly affects forecasting success.
Factor model fine if ‘driving forces’ primarily factors, so a few factors account for most variance.
Ideal case for factor model forecasting is when DGP is:

$$\mathbf{x}_t = \boldsymbol{\Upsilon}(L)\,\mathbf{f}_t + \mathbf{e}_t$$

$$\mathbf{f}_t = \boldsymbol{\Phi}(L)\,\mathbf{f}_{t-1} + \boldsymbol{\eta}_t$$

where $\mathbf{x}_t$ is $n \times 1$, $\mathbf{f}_t$ is $m \times 1$, $\boldsymbol{\Upsilon}(L)$ and $\boldsymbol{\Phi}(L)$ are $n \times m$ and $m \times m$, and $n \gg m$ so low-dimensional $\mathbf{f}_t$ drives co-movements of high-dimensional $\mathbf{x}_t$.
When mean-zero ‘idiosyncratic’ errors $\mathbf{e}_t$ satisfy $\mathsf{E}[e_{i,t}e_{j,t-k}] = 0\ i \neq j$, and $\mathsf{E}[\boldsymbol{\eta}_t\mathbf{e}_{t-k}] = \mathbf{0}$ for all k, then each $x_{i,t}$ can be optimally forecast using $\mathbf{f}_t$ and lags of $x_{i,t}$.

Factor VAR

Let $\boldsymbol{\lambda}_i(L)'$ denote i^{th} row of $\boldsymbol{\Upsilon}(L)$, then ignoring parameter estimation uncertainty:

$$
\begin{aligned}
&\mathsf{E}_t\left[x_{i,t+1} \mid \mathbf{x}_t, \mathbf{f}_t, \mathbf{x}_{t-1}, \mathbf{f}_{t-1}, \ldots\right] \\
&\quad = \mathsf{E}_t\left[\boldsymbol{\lambda}_i\left(L\right)' \mathbf{f}_{t+1} \mid \mathbf{f}_t, \mathbf{f}_{t-1}, \ldots\right] + \mathsf{E}_t\left[e_{i,t+1} \mid e_{i,t}, e_{i,t-1} \ldots\right] \\
&\quad = \boldsymbol{\alpha}\left(L\right)' \mathbf{f}_t + \delta\left(L\right) x_{i,t}
\end{aligned}
$$

Absent structural breaks, delivers best forecasts.
But factor structures may not be good representation of macroeconomy:
driving forces may be variables, but factor approximations may parsimoniously capture linear combinations of effects.

Needing sparsity–or not?

Selection with N (variables) $> T$ (observations) seems to hinge on sparsity of parameter vector, $\boldsymbol{\beta}$, in local DGP:

- must have fewer non-zero **functions** of parameters than T, otherwise LDGP is not identifiable.

Factor models are 'solution' to dimension constraints.
When selecting with both principal components and variables, sparsity is **not** required: $\boldsymbol{\beta}$ could contain all non-zero elements, yet fewer regressors than T.

Automatic selection handles $N \gg T$ & perfect collinearity, so no binding dimensionality constraint:
include all variables & principal components to retain both key individual determinants and factors.

Our approach

Linear models selected in-sample with IIS from:
(a) a large set of variables,
(b) over those variables' principal components (PCs), and
(c) over a candidate set including both.

Initial model has $N > T$,
and in case (c), is perfectly collinear.

Route map

(1) **Setting the scene**

(2) **Relating variables $\{\mathbf{z}_t\}$ to factors $\{\mathbf{f}_t\}$**

(3) Comparing variable-based and factor-based models

(4) Taxonomy of forecast errors in factor models

(5) Forecast failure in equilibrium-correction models

(6) Model selection with both factors and variables

(7) Forecasting US GDP and inflation

Conclusions

Relating external variables to factors

Consider $n < T$ weakly stationary $\{\mathbf{z}_t\}$ over $t = 1, \ldots, T$ generated by VAR(1) with intercept $\boldsymbol{\pi}$:

$$\mathbf{z}_t = \boldsymbol{\pi} + \boldsymbol{\Pi}\mathbf{z}_{t-1} + \mathbf{v}_t \text{ where } \mathbf{v}_t \sim \mathsf{IN}_n\left[\mathbf{0}, \boldsymbol{\Omega}_{\mathbf{v}}\right] \quad (1)$$

all eigenvalues of $\boldsymbol{\Pi}$ inside unit circle:

$$\mathsf{E}\left[\mathbf{z}_t\right] = \boldsymbol{\pi} + \boldsymbol{\Pi}\mathsf{E}\left[\mathbf{z}_{t-1}\right] = \boldsymbol{\pi} + \boldsymbol{\Pi}\boldsymbol{\mu} = \boldsymbol{\mu}$$

where $\boldsymbol{\mu} = (\mathbf{I}_n - \boldsymbol{\Pi})^{-1}\boldsymbol{\pi}$.

Principal-component description is:

$$\mathbf{z}_t = \boldsymbol{\Psi}\mathbf{f}_t + \mathbf{e}_t \text{ with } \mathbf{e}_t \sim \mathsf{ID}_n\left[\mathbf{0}, \boldsymbol{\Omega}_{\mathbf{e}}\right] \quad (2)$$

$\mathbf{f}_t \sim \mathsf{ID}_m[\boldsymbol{\kappa}, \mathbf{P}]$ of dimension $m \leq n$, $\boldsymbol{\Psi}$ is $n \times m$, $\mathsf{E}[\mathbf{f}_t\mathbf{e}_t'] = \mathbf{0}$.

When $\mathsf{E}[\mathbf{f}_t] = \boldsymbol{\kappa}$ and $\mathsf{E}[\mathbf{e}_t] = \mathbf{0}$, from (2):

$$\mathsf{E}\left[\mathbf{z}_t\right] = \boldsymbol{\Psi}\mathsf{E}\left[\mathbf{f}_t\right] + \mathsf{E}\left[\mathbf{e}_t\right] = \boldsymbol{\Psi}\boldsymbol{\kappa} = \boldsymbol{\mu} \quad (3)$$

Eigenvector decompositions

Then:

$$\mathsf{E}\left[(\mathbf{z}_t-\boldsymbol{\mu})(\mathbf{z}_t-\boldsymbol{\mu})'\right]=\mathbf{\Psi}\mathsf{E}\left[(\mathbf{f}_t-\boldsymbol{\kappa})(\mathbf{f}_t-\boldsymbol{\kappa})'\right]\mathbf{\Psi}'+\mathsf{E}\left[\mathbf{e}_t\mathbf{e}_t'\right]$$
$$=\mathbf{\Psi}\mathbf{P}\mathbf{\Psi}'+\mathbf{\Omega}_{\mathbf{e}}=\mathbf{M} \quad (4)$$

where $\mathbf{P}$ is $m\times m$ diagonal and hence $\mathbf{z}_t\sim\mathsf{D}_n[\boldsymbol{\mu},\mathbf{M}]$. Let:

$$\mathbf{M}=\mathbf{B}\mathbf{\Lambda}\mathbf{B}' \quad (5)$$

where $\mathbf{B}'\mathbf{B}=\mathbf{I}_n$, so $\mathbf{B}^{-1}=\mathbf{B}'$ with eigenvalues ordered from largest downwards:

$$\mathbf{B}'=\begin{pmatrix}\mathbf{B}_1'\\ \mathbf{B}_2'\end{pmatrix} \text{ and } \mathbf{\Lambda}=\begin{pmatrix}\mathbf{\Lambda}_{11} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Lambda}_{22}\end{pmatrix}, \quad (6)$$

where $\mathbf{\Lambda}_{11}$ is $m\times m$, with $\mathbf{B}_1'\mathbf{M}\mathbf{B}_1=\mathbf{\Lambda}_{11}$ and:

$$\mathbf{B}\mathbf{\Lambda}\mathbf{B}'=\mathbf{B}_1\mathbf{\Lambda}_{11}\mathbf{B}_1'+\mathbf{B}_2\mathbf{\Lambda}_{22}\mathbf{B}_2'.$$

Factors

Consequently, from (2) and (6):

$$\mathbf{B}'(\mathbf{z}_t - \boldsymbol{\mu}) = \mathbf{B}'(\boldsymbol{\Psi}(\mathbf{f}_t - \boldsymbol{\kappa}) + \mathbf{e}_t) = \mathbf{f}_t - \boldsymbol{\kappa} \tag{7}$$

If only m actually matter, $\mathbf{B}_1'$ weights the $\mathbf{z}_t$ to produce $\mathbf{f}_{1,t}$ where:

$$\mathbf{B}_1'(\mathbf{z}_t - \boldsymbol{\mu}) = \mathbf{f}_{1,t} - \boldsymbol{\kappa}_1 \tag{8}$$

where only $\mathbf{f}_{1,t}$ are relevant to explaining y_t.

Simulating factors and variables

When no factors matter, DGP over $t = 1, \ldots, T$ is:

$$y_t = \boldsymbol{\beta}'\mathbf{z}_t + \epsilon_t \quad (9)$$

where $\epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right]$. Regressors generated by:

$$\mathbf{z}_t \sim \mathsf{IN}_n\left[\mathbf{0}, \sigma^2\boldsymbol{\Omega}\right], \quad t = 1, \ldots, T \quad (10)$$

where $\boldsymbol{\Omega} = (1-\rho)\mathbf{I}_n + \rho\boldsymbol{\iota}\boldsymbol{\iota}'$, $T = 100$, with $\rho = 0.5, 0.9$, $n = 2, 10, 50$ using $M = 10,000$ replications in *Autometrics* at $\alpha = 0.05, 0.01$, with no diagnostic checking.

Most general GUM is given by:

$$y_t = \boldsymbol{\gamma}'\mathbf{z}_t + \boldsymbol{\delta}'\mathbf{f}_t + \nu_t \quad (11)$$

where $\boldsymbol{\delta} = \mathbf{0}$, leading to $2n$ regressors, with n collinearities.

Table 1 records the gauge: close to α when no collinearity, well under with collinearity, till $n = 50$, when expanding and contracting block searches return $g \simeq \alpha$.

Simulation outcomes under the null

	$\alpha = 0.05$			$\alpha = 0.01$		
	$n = 2$	$n = 10$	$n = 50$	$n = 2$	$n = 10$	$n = 50$
$\rho = 0.5$						
(i)	0.052	0.048	0.031	0.012	0.009	0.005
(ii)	0.052	0.048	0.031	0.012	0.009	0.005
(iii)	0.024	0.002	0.047	0.004	0.0002	0.011
$\rho = 0.9$						
(i)	0.052	0.057	0.031	0.011	0.013	0.006
(ii)	0.052	0.057	0.031	0.011	0.013	0.006
(iii)	0.026	0.002	0.048	0.004	0.0002	0.013

Table 1: Gauge for selection with (i) just variables, (ii) just factors, and (iii) variables and factors.

Analyzing factors approximating variables

First, close to a 'factor DGP': factor structure of $\mathbf{z}_t$ in (10) matches relation between y_t and $\mathbf{z}_t$ in (9). For $n = 3$:

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix} \qquad (12)$$

which has an eigenvector of $\mathbf{h}_1 = \boldsymbol{\iota} = (1 \ldots 1)'$ so:

$$\phi f_{1,t} = \phi \boldsymbol{\iota}' \mathbf{z}_t = \boldsymbol{\beta}' \mathbf{z}_t. \qquad (13)$$

When DGP in (9) has $\phi = \boldsymbol{\iota}'\boldsymbol{\beta}$:

$$y_t = \phi f_{1,t} + \epsilon_t \qquad (14)$$

with non-centrality ψ of the t-test of $\phi = 0$:

$$\psi = \frac{\widehat{\phi}}{\mathsf{SE}\left[\widehat{\phi}\right]} \simeq \phi \frac{\sigma_f \sqrt{n\left(1 + (n-1)\rho\right)}}{\sigma_\epsilon} \qquad (15)$$

Analyzing factors approximating variables

When $\sigma_f^2 = \sigma_\epsilon^2 = 1$, $\rho = 0.9$ and $n = 3$, ratio of factor to variable non-centralities is 7.5.
(Ratios of 28.8 at $n = 10$ and 162 at $n = 50$)
When many small effects, in this setting PCs retained with high probability relative to retention of individual regressors.

Huge advantage to representing individually-insignificant effects in $\mathbf{z}_t$ by factors:
large reduction in RMSEs relative to just estimating DGP.
Bias correction downweights retained irrelevant factors, yielding yet smaller RMSEs;
first principal component highly significant,
captures most of variation in y

Deriving estimated parameters of interest

After selection, denote retained $\mathbf{z}_t$ and $\mathbf{w}_t$ by $\mathbf{z}_{r,t}$ and $\mathbf{w}_{s,t}$, with coefficients $\widetilde{\boldsymbol{\gamma}}_r$ and $\widetilde{\boldsymbol{\delta}}_s$:

$$y_t = \widetilde{\boldsymbol{\gamma}}'_r \mathbf{z}_{r,t} + \widetilde{\boldsymbol{\delta}}'_s \mathbf{w}_{s,t} + v_t \tag{16}$$

From (37) the retained principal components are:

$$\mathbf{w}_{s,t} = \widehat{\mathbf{H}}_s \widetilde{\mathbf{z}}_t$$

Solving out for the principal components results in:

$$\begin{aligned} y_t &= \widetilde{\boldsymbol{\gamma}}'_r \mathbf{z}_{r,t} + \widetilde{\boldsymbol{\delta}}'_s \mathbf{w}_{s,t} + v_t \\ &= \widetilde{\boldsymbol{\gamma}}'_{r*} \mathbf{z}_t + \widetilde{\boldsymbol{\delta}}'_{s*} \mathbf{w}_t + v_t \\ &= \lambda_s + \widetilde{\boldsymbol{\gamma}}'_s \mathbf{z}_t + v_t \end{aligned} \tag{17}$$

$\boldsymbol{\gamma}_{r*}$ and $\boldsymbol{\delta}_{s*}$ augment $\boldsymbol{\gamma}_r$ and $\boldsymbol{\delta}_s$ by zeros for non-selected elements.

Evaluating selection

First k regressors relevant, $n-k$ irrelevant.

$\widetilde{\beta}^{*}_{j,i}$ denotes OLS coefficient estimate after selection for jth regressor in replication i, with M replications.

$$\overline{\beta}^{*}_{j} = \frac{\sum_{i=1}^{M}\left[\widetilde{\beta}^{*}_{j,i} 1\left(\widetilde{\beta}^{*}_{j,i} \neq 0\right)\right]}{\sum_{i=1}^{M} 1\left(\widetilde{\beta}^{*}_{j,i} \neq 0\right)}, \quad j = 1,\ldots,n,$$

$$\text{bias}_{j} = \frac{\sum_{i=1}^{M}\left[\left(\widetilde{\beta}^{*}_{j,i} - \beta_{j}\right) 1\left(\widetilde{\beta}^{*}_{j,i} \neq 0\right)\right]}{\sum_{i=1}^{M} 1\left(\widetilde{\beta}^{*}_{j,i} \neq 0\right)}, \quad j = 1,\ldots,n$$

$$\text{RMSE}_{j} = \left[\frac{\left[\sum_{i=1}^{M}\left(\widetilde{\beta}^{*}_{j,i} - \beta_{j}\right)^{2} 1\left(\widetilde{\beta}^{*}_{j,i} \neq 0\right)\right]}{\sum_{i=1}^{M} 1\left(\widetilde{\beta}^{*}_{j,i} \neq 0\right)}\right]^{1/2}, \quad j = 1,\ldots,n$$

or β^{2}_{j} when $\sum_{i=1}^{M} 1(\widetilde{\beta}^{*}_{j,i} \neq 0) = 0$.

Correctly approximating small effects by PCs

	[A]	[B] 5%	[C] 5%	[B] 1%	[C] 1%	[D]	[E] 5%	[F] 5%	[E] 1%	[F] 1%
$n = 10; \psi_i = 1; \rho = 0.5$										
$\widehat{\sigma}$	0.998	0.995	0.995	1.012	1.012	1.004	0.993	0.993	1.001	1.001
$\overline{n}$							1.443	1.443	1.091	1.091
Bias	-0.022	23.54	17.13	31.56	25.50	-0.023	-0.069	-0.075	-0.082	-0.083
RMSE	**0.143**	0.253	0.217	0.330	0.290	0.144	0.076	0.052	0.044	**0.032**
$n = 10; \psi_i = 1; \rho = 0.9$										
$\widehat{\sigma}$	0.998	0.994	0.994	1.010	1.010	1.046	1.033	1.033	1.039	1.039
$\overline{n}$							1.375	1.375	1.070	1.070
Bias	-0.018	53.15	40.26	71.86	61.05	-0.018	-0.023	-0.023	-0.024	-0.024
RMSE	**0.318**	0.574	0.503	0.752	0.682	0.320	0.157	0.102	0.083	**0.056**

[A] estimating DGP; [B] selection from DGP by *Autometrics*;
[C] bias correction in [B]; [D] estimation of factor model;
[E] 1-cut selection from factor model; [F] bias correction of [E].

Poorly approximating small effects by PCs

If link of y_t to $\mathbf{z}_t$ in (9) does **not** match correlations in (10), (e.g., β_i alternate as $\pm\phi$) need all 3 factors.
Other two eigenvectors of (12) are:

$$\mathbf{h}_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{h}_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

so $\boldsymbol{\beta}'\mathbf{z}_t = \phi\begin{pmatrix} 1 & -1 & 1 \end{pmatrix}\mathbf{z}_t$ equals $\mathbf{h}_1 + 2\mathbf{h}_2 - 4\mathbf{h}_3$.

$\beta_j = -a \times -1^j$, for $j = 1, \ldots, n$ where a calibrated for $|\mathsf{t}| = 1$

A final case is where a subset of the $\mathbf{z}_t$ are irrelevant so $\boldsymbol{\beta}' = \phi\begin{pmatrix} 0 & 1 & 1 \end{pmatrix}$ or $\phi\begin{pmatrix} 1 & -1 & 0 \end{pmatrix}$. The former, e.g., is $\phi\left(2\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3\right)$ (again needing all three factors), the latter is $\phi\mathbf{h}_2$.

Simulating poor small-effects approximation by PCs

	[A]	[B]	[C]	[D]	[E]	[F]
		1%	1%		1%	1%
$n = 10;\ \psi_i = (-1)^i;\ \rho = 0.5$						
$\widehat{\sigma}$	0.998	1.021	1.021	0.999	0.999	0.999
$\overline{n}$					0.636	0.636
Bias $(+)$	-0.020	22.02	12.63	-0.016	-5.470	-7.411
Bias $(-)$	-0.025	-21.92	-12.56	-0.029	5.464	7.406
RMSE	**0.143**	0.243	0.182	0.144	0.145	**0.130**
$n = 10;\ \psi_i = (-1)^i;\ \rho = 0.9$						
$\widehat{\sigma}$	0.998	1.016	1.016	0.999	1.000	1.000
$\overline{n}$					0.650	0.650
Bias $(+)$	0.011	46.83	33.21	0.019	-12.38	-16.78
Bias $(-)$	-0.047	-47.09	-33.37	-0.056	12.38	16.78
RMSE	**0.318**	0.522	0.420	0.320	0.325	**0.289**

[A] estimating DGP; [B] selection from DGP by *Autometrics*;
[C] bias correction in [B]; [D] estimation of factor model;
[E] 1-cut selection from factor model; [F] bias correction of [E].

Route map

(1) **Setting the scene**

(2) **Relating variables $\{\mathbf{z}_t\}$ to factors $\{\mathbf{f}_t\}$**

(3) **Comparing variable-based and factor-based models**

(4) Taxonomy of forecast errors in factor models

(5) Forecast failure in equilibrium-correction models

(6) Model selection with both factors and variables

(7) Forecasting US GDP and inflation

Conclusions

Variable-based and factor-based models

The in-sample DGP for y_t is:

$$y_t = \beta_0 + \boldsymbol{\beta}'\mathbf{z}_{t-1} + \rho y_{t-1} + \epsilon_t \tag{18}$$

where $|\rho| < 1$ and $\epsilon_t \sim \mathsf{IN}[0, \sigma^2_\epsilon]$, so:

$$\mathsf{E}\left[y_t\right] = \beta_0 + \boldsymbol{\beta}'\mathsf{E}\left[\mathbf{z}_{t-1}\right] + \rho\mathsf{E}\left[y_{t-1}\right] = \beta_0 + \boldsymbol{\beta}'\boldsymbol{\mu} + \rho\delta = \delta \tag{19}$$

where $\delta = (\beta_0 + \boldsymbol{\beta}'\boldsymbol{\mu})/(1-\rho)$ so (18) is the EqCM:

$$y_t - \delta = \boldsymbol{\beta}'\left(\mathbf{z}_{t-1} - \boldsymbol{\mu}\right) + \rho\left(y_{t-1} - \delta\right) + \epsilon_t \tag{20}$$

Only the subset $\mathbf{z}_{a,t-1}$ matter substantively:

$$y_t - \delta \simeq \boldsymbol{\beta}'_a\left(\mathbf{z}_{a,t-1} - \boldsymbol{\mu}_a\right) + \rho_a\left(y_{t-1} - \delta\right) + u_t \tag{21}$$

Alternatively, combinations $\mathbf{f}_{1,t}$ might matter:

$$y_t - \delta = \boldsymbol{\beta}'\mathbf{B}\left(\mathbf{f}_{t-1} - \boldsymbol{\kappa}\right) + \rho\left(y_{t-1} - \delta\right) + \epsilon_t \tag{22}$$

so that:

$$y_t - \delta = \boldsymbol{\tau}'_1\left(\mathbf{f}_{1,t-1} - \boldsymbol{\kappa}_1\right) + \rho_1\left(y_{t-1} - \delta\right) + w_t \tag{23}$$

Both variables and factors

When both variables and PCs are allowed:

$$y_t - \delta = \boldsymbol{\beta}_b' \left(\mathbf{z}_{b,t-1} - \boldsymbol{\mu}_b\right) + \boldsymbol{\tau}_2' \left(\mathbf{f}_{2,t-1} - \boldsymbol{\kappa}_2\right) + \rho_b \left(y_{t-1} - \delta\right) + \epsilon_t \quad (24)$$

Use IIS to handle location shifts and outliers in-sample.

If $2n$ variables and factors at significance level α, $2\alpha n$ will be adventitiously significant on average– **one** retained by chance when $n = 100$ and $\alpha = 0.005$: critical value of $c_\alpha \simeq 3.10$.

When DGP for y_t is an AR(1):

$$y_t = \gamma_0 + \gamma_1 y_{t-1} + v_t.$$

If search **only** over factors, retain many $\mathbf{f}_t$ to approximate y_{t-1} so starting model always includes y_{t-1}.

Forecasting from variables and factors

Forecast $\{y_{T+h}\}$ over horizon $h = 1, \ldots, H$,
from forecast origin at T, with information set $\mathbf{Z}_T^1 = (\mathbf{z}_1 \ldots \mathbf{z}_T)$ and $(y_1 \ldots y_T)$.

For in-sample estimates $\{\widehat{\mathbf{f}}_t\}$, 1-step forecasts are:

$$\widehat{y}_{T+1|T} = \widehat{\delta} + \widehat{\boldsymbol{\beta}}_b' \left(\mathbf{z}_{b,T} - \widehat{\boldsymbol{\mu}}_b\right) + \widehat{\boldsymbol{\tau}}_2' \left(\widehat{\mathbf{f}}_{2,T} - \widehat{\boldsymbol{\kappa}}_2\right) + \widehat{\rho} \left(\widehat{y}_T - \widehat{\delta}\right) \quad (25)$$

where $\widehat{y}_T$ is the ‘flash’ estimate.
Use multi-step estimation for h-step ahead forecasts:

$$\widehat{y}_{T+h|T} = \widehat{\delta}_{(h)} + \widehat{\boldsymbol{\beta}}_{b,(h)}' \left(\mathbf{z}_{b,T} - \widehat{\boldsymbol{\mu}}_b\right) + \widehat{\boldsymbol{\tau}}_{2,(h)}' \left(\widehat{\mathbf{f}}_{2,T} - \widehat{\boldsymbol{\kappa}}_2\right) + \widehat{\rho}_{(h)} \left(\widehat{y}_T - \widehat{\delta}_{(h)}\right) \quad (26)$$

so $\widehat{u}_{T+h|T} = y_{T+h} - \widehat{y}_{T+h|T}$, a moving average of order $(h-1)$.

Route map

(1) **Setting the scene**

(2) **Relating variables $\{\mathbf{z}_t\}$ to factors $\{\mathbf{f}_t\}$**

(3) **Comparing variable-based and factor-based models**

(4) **Taxonomy of forecast errors in factor models**

(5) Forecast failure in equilibrium-correction models

(6) Model selection with both factors and variables

(7) Forecasting US GDP and inflation

Conclusions

Location shifts

DGP depends on $\mathbf{z}_{t-1}$ and y_{t-1},
forecasting model incorporates a subset of the factors.

Highlight principle sources of
forecast bias and forecast-error variance
after location shifts where DGP changes at T to:

$$y_{T+h} = \delta^* + \boldsymbol{\beta}' \left(\mathbf{z}_{T+h-1} - \boldsymbol{\mu}^*\right) + \rho \left(y_{T+h-1} - \delta^*\right) + \epsilon_{T+h} \qquad (27)$$

for $h = 1, \ldots, H$.

Mapping to principal components yields:

$$y_{T+h} = \delta^* + \boldsymbol{\tau}' \left(\mathbf{f}_{T+h-1} - \boldsymbol{\kappa}^*\right) + \rho \left(y_{T+h-1} - \delta^*\right) + \epsilon_{T+h} \qquad (28)$$

where for now $\boldsymbol{\tau}$ and ρ remain at in-sample values during forecast period.

Taxonomy of forecast errors

1-step forecast-error taxonomy distinguishes 11 sources.

From (28) minus (25), for $h = 1$:

$$\widehat{u}_{T+1|T} = \left(\delta^* - \widehat{\delta}\right) + \boldsymbol{\tau}'\left(\mathbf{f}_T - \boldsymbol{\kappa}^*\right) - \widehat{\boldsymbol{\tau}}_1'\left(\widehat{\mathbf{f}}_{1,T} - \widehat{\boldsymbol{\kappa}}_1\right)$$
$$+ \rho\left(y_T - \delta^*\right) - \widehat{\rho}\left(\widehat{y}_T - \widehat{\delta}\right) + \epsilon_{T+1}.$$

Using $\boldsymbol{\tau}_1'(\boldsymbol{\kappa}_1^* - \boldsymbol{\kappa}_1) + \boldsymbol{\tau}_2'(\boldsymbol{\kappa}_2^* - \boldsymbol{\kappa}_2) = \boldsymbol{\tau}'(\boldsymbol{\kappa}^* - \boldsymbol{\kappa})$, yields table 2.

Factor model taxonomy of forecast errors

Table 2: $\widehat{u}_{T+1|T} = \ldots$

$(1-\rho)(\delta^*-\delta)$	[A] equilibrium-mean shift
$-\boldsymbol{\tau}'(\boldsymbol{\kappa}^*-\boldsymbol{\kappa})$	[B] factor-mean shift
$+(1-\rho)\left(\delta-\widehat{\delta}\right)$	[C] equilibrium-mean estimation
$-\boldsymbol{\tau}_1'(\boldsymbol{\kappa}_1-\widehat{\boldsymbol{\kappa}}_1)$	[D] factor-mean estimation
$+\rho(y_T-\widehat{y}_T)$	[E] flash estimate error
$+\boldsymbol{\tau}_1'\left(\mathbf{f}_{1,T}-\widehat{\mathbf{f}}_{1,T}\right)$	[F] factor estimate error
$+\boldsymbol{\tau}_2'(\mathbf{f}_{2,T}-\boldsymbol{\kappa}_2)$	[G] factor approximation error
$+(\boldsymbol{\tau}_1-\widehat{\boldsymbol{\tau}}_1)'\left(\widehat{\mathbf{f}}_{1,T}-\boldsymbol{\kappa}_1\right)$	[H] factor estimation covariance
$+(\rho-\widehat{\rho})\left(\widehat{y}_T-\widehat{\delta}\right)$	[I] flash estimation covariance
$+(\boldsymbol{\tau}_1-\widehat{\boldsymbol{\tau}}_1)'(\boldsymbol{\kappa}_1-\widehat{\boldsymbol{\kappa}}_1)$	[J] parameter estimation covariance
$+\epsilon_{T+1}$	[K] innovation error

Interpretation

If unbiased estimates, neglecting terms of $o_p(T^{-1})$:

$$\mathsf{E}\left[\widehat{u}_{T+1|T}\right] \simeq (1-\rho)\left(\delta^* - \delta\right) - \boldsymbol{\tau}'\left(\boldsymbol{\kappa}^* - \boldsymbol{\kappa}\right) + \rho\left(y_T - \mathsf{E}[\widehat{y}_T]\right) + \boldsymbol{\tau}_1'\left(\mathbf{f}_{1,T} - \mathsf{E}[\widehat{\mathbf{f}}_{1,T}]\right) \qquad (29)$$

[A] and [B] primary determinants of forecast bias:
data and factor-estimation errors [E] & [F] contribute.
Those & remaining terms add to forecast-error variance.
Factor approximation error not in (29) as $\mathsf{E}[\mathbf{f}_{2,T}] = \boldsymbol{\kappa}_2$.
Even if [E] and [F] negligible, shifts could be large, if in (1):

$$\boldsymbol{\pi}^* = \boldsymbol{\pi} + 1_{(t \geq T)}\boldsymbol{\theta} \;\text{ for }\; h = 1, \ldots, H \qquad (30)$$

so permanently shifts at T:

$$\boldsymbol{\pi} = \left(\mathbf{I}_n - \boldsymbol{\Pi}\right)\boldsymbol{\Psi}\boldsymbol{\kappa}$$

When $\boldsymbol{\Pi}$ and $\boldsymbol{\Psi}$ are constant, $\boldsymbol{\kappa}$ will shift, so for $n = m$:

$$\boldsymbol{\kappa}^* = \boldsymbol{\Psi}^{-1}\left(\mathbf{I}_n - \boldsymbol{\Pi}\right)^{-1}\boldsymbol{\pi}^* = \boldsymbol{\kappa} + 1_{(t \geq T)}\boldsymbol{\Psi}^{-1}\left(\mathbf{I}_n - \boldsymbol{\Pi}\right)^{-1}\boldsymbol{\theta} \qquad (31)$$

Slope shifts

Forecast-error biases entailed by equilibrium-mean shifts within model of y_{T+1} ($\delta^* \neq \delta$) or in external variables ($\boldsymbol{\kappa}^* \neq \boldsymbol{\kappa}$) irrespective of inclusion or exclusion of associated factors: approximation error does not induce such a problem.

If $\boldsymbol{\tau}$ and ρ change in the forecast period:

$$y_{T+1} = \delta^* + \boldsymbol{\tau}^{*\prime} \left(\mathbf{f}_T - \boldsymbol{\kappa}^* \right) + \rho^* \left(y_T - \delta^* \right) + \epsilon_{T+1} \qquad (32)$$

the change in $\boldsymbol{\tau}$ induces an additional error term:

$$\boldsymbol{\tau}^{*\prime} \left(\mathbf{f}_T - \boldsymbol{\kappa}^* \right) - \boldsymbol{\tau}' \left(\mathbf{f}_T - \boldsymbol{\kappa}^* \right) = \left(\boldsymbol{\tau}^{*\prime} - \boldsymbol{\tau}' \right) \left(\mathbf{f}_T - \boldsymbol{\kappa}^* \right)$$

interacting with the location shift, but otherwise is benign– this additional term will not contribute to bias when $\boldsymbol{\kappa}^* = \boldsymbol{\kappa}$ Similarly, the shift in ρ only affects forecast bias if δ also changes via:

$$\left(\rho^* - \rho \right) \left(y_T - \delta^* \right)$$

which has a zero expectation when $\delta^* = \delta$.

AR(1) taxonomy

Principal sources of forecast error for an AR(1) model ignoring parameter estimation uncertainty and data mis-measurement:

$$y_t = \delta + \gamma\left(y_{t-1} - \delta\right) + v_t \tag{33}$$

when the forecast period DGP is (28).

Omitting factors will alter γ, so $\gamma \neq \rho$, but not affect δ.

Forecast error from the AR(1) model is $\widehat{v}_{T+1|T}$, so:

$$\widehat{v}_{T+1|T} = (1-\rho)\left(\delta^* - \delta\right) - \boldsymbol{\tau}'\left(\boldsymbol{\kappa}^* - \mathbf{f}_T\right) + (\rho - \gamma)\left(y_T - \delta\right)$$

with a forecast bias of:

$$\mathsf{E}\left[\widehat{v}_{T+1|T}\right] = (1-\rho)\left(\delta^* - \delta\right) - \boldsymbol{\tau}'\left(\boldsymbol{\kappa}^* - \boldsymbol{\kappa}\right)$$

matching the two leading terms in (29) for the bias of the factor-forecasting model.

Including 'correct' factors, a subset, or none has no effect on forecast biases following a location shift.

Route map

(1) **Setting the scene**

(2) **Relating variables $\{\mathbf{z}_t\}$ to factors $\{\mathbf{f}_t\}$**

(3) **Comparing variable-based and factor-based models**

(4) **Taxonomy of forecast errors in factor models**

(5) **Forecast failure in equilibrium-correction models**

(6) Model selection with both factors and variables

(7) Forecasting US GDP and inflation

Conclusions

The equilibrium-correction problem

Forecasting is not one-off venture, so compare models as the origin moves through time.
All models will fail when there is an unknown location shift but speed and extent to which forecasts recover differs.
'Equilibrium-correction' models, to which (25) belongs, lack adaptability over time.
Write (25) for 1-step forecasts:

$$\Delta\widehat{y}_{T+1|T} = \widehat{\boldsymbol{\tau}}_1' \left(\widehat{\mathbf{f}}_{1,T} - \widehat{\boldsymbol{\kappa}}_1\right) + \left(\widehat{\rho} - 1\right)\left(\widehat{y}_T - \widehat{\delta}\right)$$

so that $\mathsf{E}[\Delta\widehat{y}_{T+1|T}] \simeq 0$, whereas the DGP is:

$$\Delta y_{T+1} = \boldsymbol{\tau}'\left(\mathbf{f}_T - \boldsymbol{\kappa}^*\right) + \left(\rho - 1\right)\left(y_T - \delta^*\right) + \epsilon_{T+1} \qquad (34)$$

expected value is non-zero when there are locations shifts:

$$\begin{aligned} \mathsf{E}\left[\Delta y_{T+1}\right] &= \boldsymbol{\tau}'\mathsf{E}\left[\mathbf{f}_T - \boldsymbol{\kappa}^*\right] + \left(\rho - 1\right)\mathsf{E}\left[y_T - \delta^*\right] \\ &= \boldsymbol{\tau}_1'\left(\boldsymbol{\kappa}_1 - \boldsymbol{\kappa}_1^*\right) + \left(\rho - 1\right)\left(\delta - \delta^*\right) \qquad (35) \end{aligned}$$

Correcting equilibrium corrections

In EqCMs, problem persists: forecasting $T+2$ from $T+1$ even for known in-sample parameters, etc:

$$\begin{aligned}\Delta\widehat{u}_{T+2|T+1} &= \Delta y_{T+2} - \Delta\widehat{y}_{T+2|T+1} \\ &= \boldsymbol{\tau}_1'\left(\boldsymbol{\kappa}_1 - \boldsymbol{\kappa}_1^*\right) + (\rho-1)\left(\delta-\delta^*\right) + \epsilon_{T+1}\end{aligned}$$

Use a robust forecasting device exploiting (35):

$$\Delta\widetilde{y}_{T+2|T+1} = \Delta y_{T+1} + \widehat{\boldsymbol{\tau}}_1'\Delta\widehat{\mathbf{f}}_{1,T} + (\widehat{\rho}-1)\,\Delta\widehat{y}_T$$

Again under simplifying assumptions, forecast error $\Delta\widetilde{u}_{T+2|T+1} = \Delta y_{T+2} - \Delta\widetilde{y}_{T+2|T+1}$ is:

$$\begin{aligned}\Delta\widetilde{u}_{T+2|T+1} &= \boldsymbol{\tau}'\left(\mathbf{f}_{T+1} - \boldsymbol{\kappa}^*\right) + (\rho-1)\left(y_{T+1}-\delta^*\right) + \epsilon_{T+2} \\ &\quad - \boldsymbol{\tau}'\left(\mathbf{f}_T - \boldsymbol{\kappa}^*\right) - (\rho-1)\left(y_T-\delta^*\right) - \epsilon_{T+1} \\ &= \boldsymbol{\tau}_1'\Delta\mathbf{f}_{1,T+1} + (\rho-1)\,\Delta y_{T+1} + \Delta\epsilon_{T+2} \qquad (36)\end{aligned}$$

so less dependent on location shifts.

Forecasting Japanese exports

Illustrate by forecasting Japanese exports over 2008(7)–2011(6)

Model selected at 1% by Autometrics, 2000(1)–2008(6):

$$\widehat{y}_t = \underset{(0.091)}{0.68}\, y_{t-1} + \underset{(0.090)}{0.26}\, y_{t-3} + \underset{(0.04)}{0.12}\, 1_{2000(2)} + \underset{(0.04)}{0.12}\, 1_{2002(1)}$$

$$\widehat{\sigma} = 0.039 \quad \chi^2(2) = 0.65$$

$$\mathsf{F}_{\mathsf{het}}(4,95) = 1.02 \quad \mathsf{F}_{\mathsf{ar}}(6,92) = 0.97 \quad \mathsf{F}_{\mathsf{reset}}(2,96) = 2.1$$

$1_{200z(x)}$ indicators for $200z(x)$.

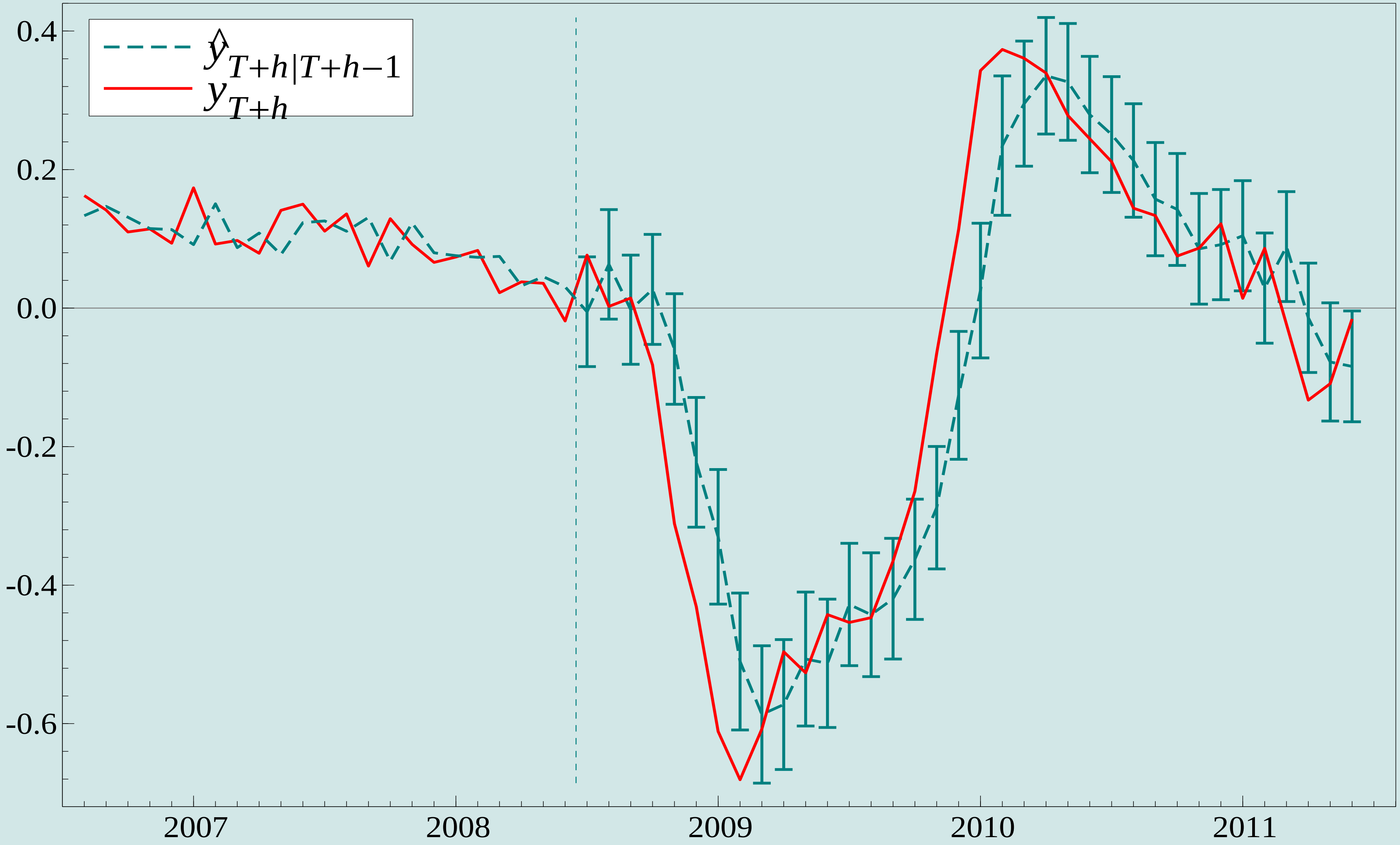
Almost ‘robust’ by near second unit root.

The corresponding robust device, therefore, was:

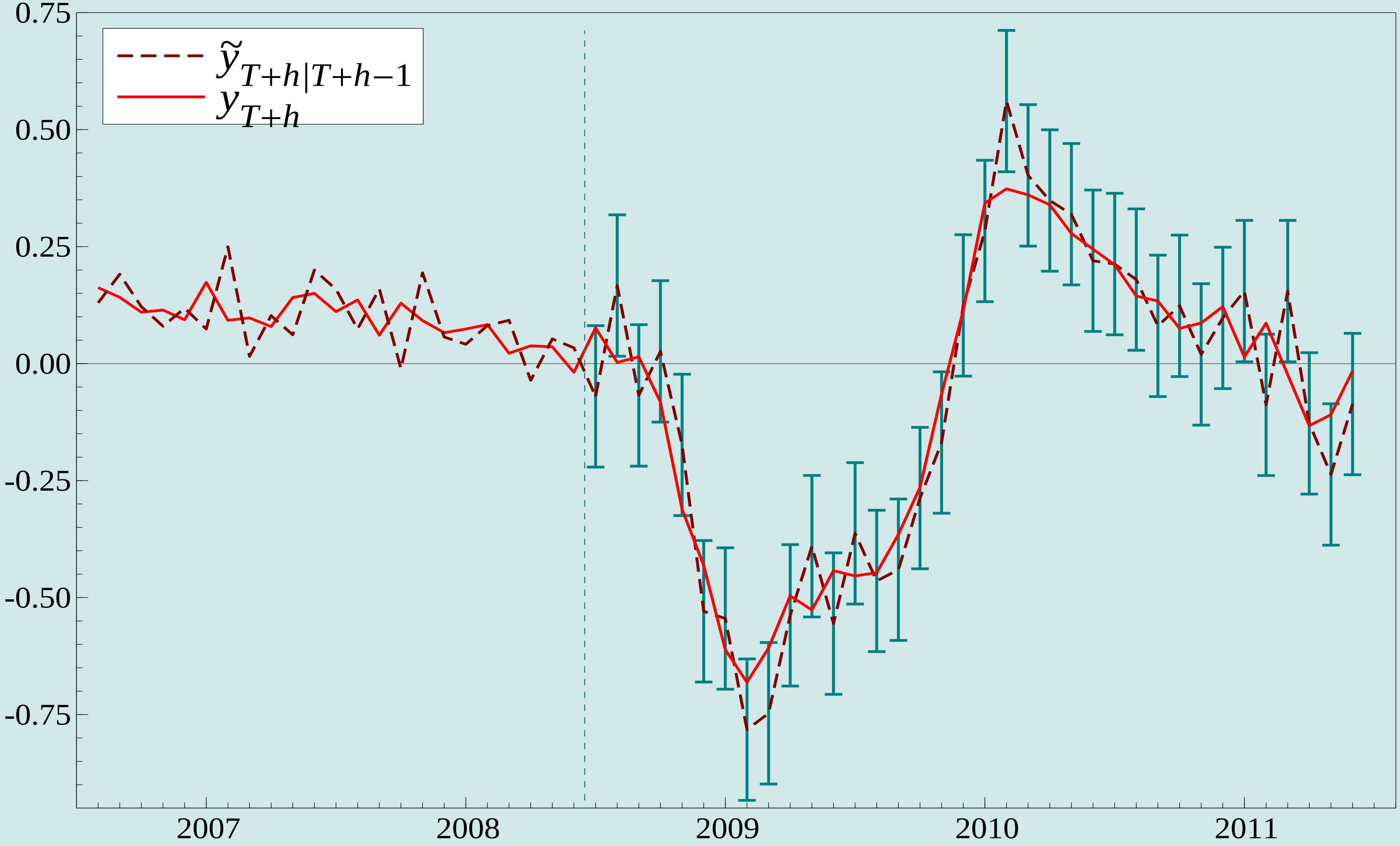
$$\widetilde{y}_t = y_{t-1} + 0.95\Delta y_{t-1} \qquad \widetilde{\sigma} = 0.0755$$

Their respective forecasts follow.

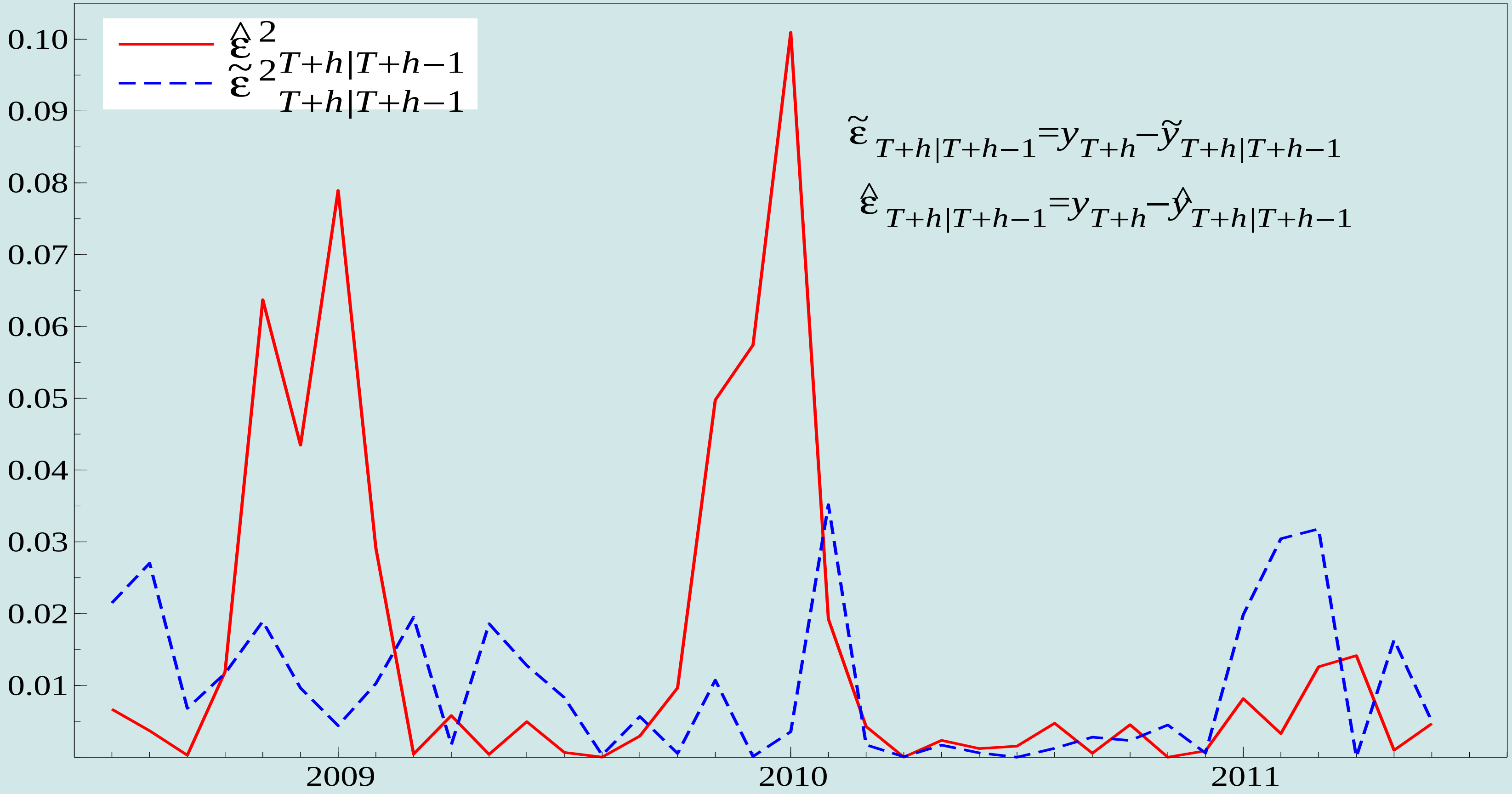
Model-based forecasts 2008(7)–2011(6)



Robust forecasts 2008(7)–2011(6)



Squared forecast-error comparison



Typical pattern: robust device avoids forecast failure, at an insurance cost when no shifts occur.
Overall RMSFEs of 0.124 versus 0.098.

Factor models

When factor models are EqCMs, location shifts:
(a) affect the mapping between the $\mathbf{z}_{T+1}$ and $\mathbf{f}_{T+1}$ by changes in the weights:
Stock and Watson (2009) find a relatively innocuous effect;
(b) breaks in coefficients of zero-mean variables or factors not problematic;
(c) location shifts induce systematic mis-forecasting.

In their forecasting exercise, variables differenced once, so large equilibrium-mean shifts unlikely:
partially robust to previous location, but not growth, shifts.

Robust forecasts

If the differenced-data version of (33) is used, forecasting $T+2$ from $T+1$:

$$\Delta\widetilde{y}_{T+2|T+1} = \gamma\Delta y_{T+1}$$

when:

$$\Delta y_{T+2} = \boldsymbol{\tau}^{*\prime}\Delta\mathbf{f}_{T+1} + \rho^{*}\Delta y_{T+1} + \Delta\epsilon_{T+2}$$

we have:

$$\widetilde{v}_{T+2|T+1} = \boldsymbol{\tau}^{*\prime}\Delta\mathbf{f}_{T+1} + \left(\rho^{*} - \gamma\right)\Delta y_{T+1} + \Delta\epsilon_{T+2}$$

which is close to (36).

Route map

(1) **Setting the scene**

(2) **Relating variables $\{\mathbf{z}_t\}$ to factors $\{\mathbf{f}_t\}$**

(3) **Comparing variable-based and factor-based models**

(4) **Taxonomy of forecast errors in factor models**

(5) **Forecast failure in equilibrium-correction models**

(6) **Model selection with both factors and variables**

(7) Forecasting US GDP and inflation

Conclusions

Automatic model selection

Selection gives zero weight to insignificant variables,
PCs place small non-zero weight even if variables have low correlation with other $\mathbf{z}$s.
Selection avoids separate choice of PCs–based on covariation between them and y:
does not require correlation structure between $\mathbf{z}$s and y to be similar to within $\mathbf{z}$s.

Empirically, retained PCs tend not to be first few,
so correlation structure differs for y and disaggregates.

Multi-path search allows perfectly collinear candidate regressors and $N > T$.
Autometrics uses expanding and contracting block searches:
'sieve' continues until $N < T$ and no singularities.
Tree search then applied.

Route map

(1) **Setting the scene**

(2) **Relating variables $\{\mathbf{z}_t\}$ to factors $\{\mathbf{f}_t\}$**

(3) **Comparing variable-based and factor-based models**

(4) **Taxonomy of forecast errors in factor models**

(5) **Forecast failure in equilibrium-correction models**

(6) **Model selection with both factors and variables**

(7) **Forecasting US GDP and inflation**

Conclusions

US GDP and inflation

Based on Stock and Watson (2009).

Forecast quarterly US GDP growth, change in inflation & level forecasts for GDP and inflation: $H =$1997q1–2006q4.

144 seasonally-adjusted observations, estimation and selection over $T =$1962q3–2006q4, after lags.

$n = 109$ disaggregates are candidate $\mathbf{z}$s and used for $\mathbf{f}$s.

Original data in logs first or second differenced

Sample principal components computed as:

$$\widehat{\mathbf{F}} = \widehat{\mathbf{B}}'\widetilde{\mathbf{Z}} \tag{37}$$

where $\widehat{\mathbf{B}}$ is matrix of eigenvectors and $\widetilde{\mathbf{Z}}$ is standardized data.

Impulse-indicator saturation

- Stock and Watson (2009) identify a break in 1984: coefficients of factors in forecasting models not constant across this period.
- Factors well estimated by PCs even when loadings unstable:
  more accurate factor-model forecasts result from whole-sample factors
  but forecasting models based only on the period after 1984.
- Use full-sample PCs and full-sample selection with IIS: allows for multiple breaks and outliers jointly with selection, little efficiency loss under null of no breaks.

Forecasting models

Forecasting models selected from GUM:

$$\Delta y_t = \gamma_0 + \sum_{j=J_a}^{J_b} \gamma_j \Delta y_{t-j} + \sum_{i=1}^{n} \sum_{j=J_a}^{J_b} \beta_{i,j} \Delta z_{i,t-j} + \sum_{k=1}^{n} \sum_{j=J_a}^{J_b} \phi_{k,j} f_{k,t-j} + \sum_{l=1}^{T} \delta_l \mathbf{1}_{\{l=t\}} + \epsilon_t \quad (38)$$

(i) $\boldsymbol{\phi} = \mathbf{0}$, i.e. select over variables only;
(ii) $\boldsymbol{\beta} = \mathbf{0}$, select over factors only; and
(iii) $\boldsymbol{\phi} \neq \mathbf{0}$ and $\boldsymbol{\beta} \neq \mathbf{0}$, jointly selecting variables and factors;

(a) $\boldsymbol{\delta} = \mathbf{0}$, no IIS; and
(b) $\boldsymbol{\delta} \neq \mathbf{0}$, with IIS,

1-step: $J_a = 1$, $J_b = 4$;
4-step: $J_a = 4$, $J_b = 7$;
8-step: $J_a = 8$, $J_b = 11$.

Choice of significance level

3 benchmark forecasts:

$$\Delta\widehat{y}^{RW}_{T+k+h} = \Delta y_{T+k}$$
$$\Delta\widehat{y}^{AR(D)}_{T+k+h} = \widehat{\beta}_0 + \widehat{\beta}_1 \Delta y_{T+k}$$
$$\Delta\widehat{y}^{AR(I)}_{T+k+h} = \sum_{i=1}^{h-1} \widehat{\gamma}_0 \widehat{\gamma}_1^i + \widehat{\gamma}_1^h \Delta y_{T+k}$$

for $k = 0, \ldots, H-h$ and $h = 1, 4$ and 8.

	Variables		Factors		Both	
	No IIS	IIS	No IIS	IIS	No IIS	IIS
N	441	579	441	579	877	1015
$\alpha\%$	1 (4.4)	0.5 (3)	1 (4.4)	0.5 (3)	0.5 (4.4)	0.25 (2.5)
$\alpha\%$	0.5 (2.2)	0.1 (0.6)	0.5 (2.2)	0.1 (0.6)	0.1 (0.9)	0.05 (0.5)

Significance levels used in model selection (%); expected null retention numbers in parentheses. $T = 138$.

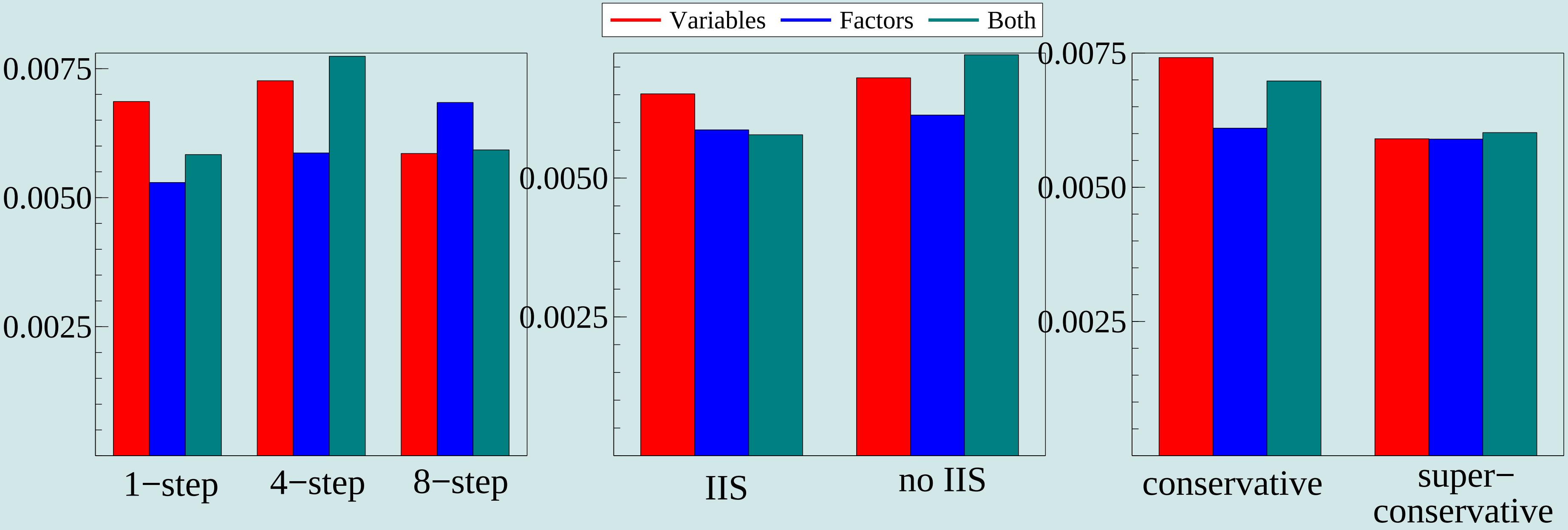
In sample fit: GDP growth with IIS

	1-step		4-step		8-step	
	Cons	Super	Cons	Super	Cons	Super
Variables						
$\widehat{\sigma}$	0.559	0.669	0.629	0.711	0.702	0.799
No. regressors	9	5	14	7	9	5
No. dummies	2	1	1	2	5	4
Factors						
$\widehat{\sigma}$	0.657	0.702	0.718	0.798	0.671	0.813
No. regressors	5	4	8	4	12	5
No. dummies	2	1	6	2	5	2
Both						
$\widehat{\sigma}$	0.553	0.753	0.712	0.819	0.767	0.788
No. regressors	9	2	9	4	9	6
No. dummies	2	0	1	0	2	4

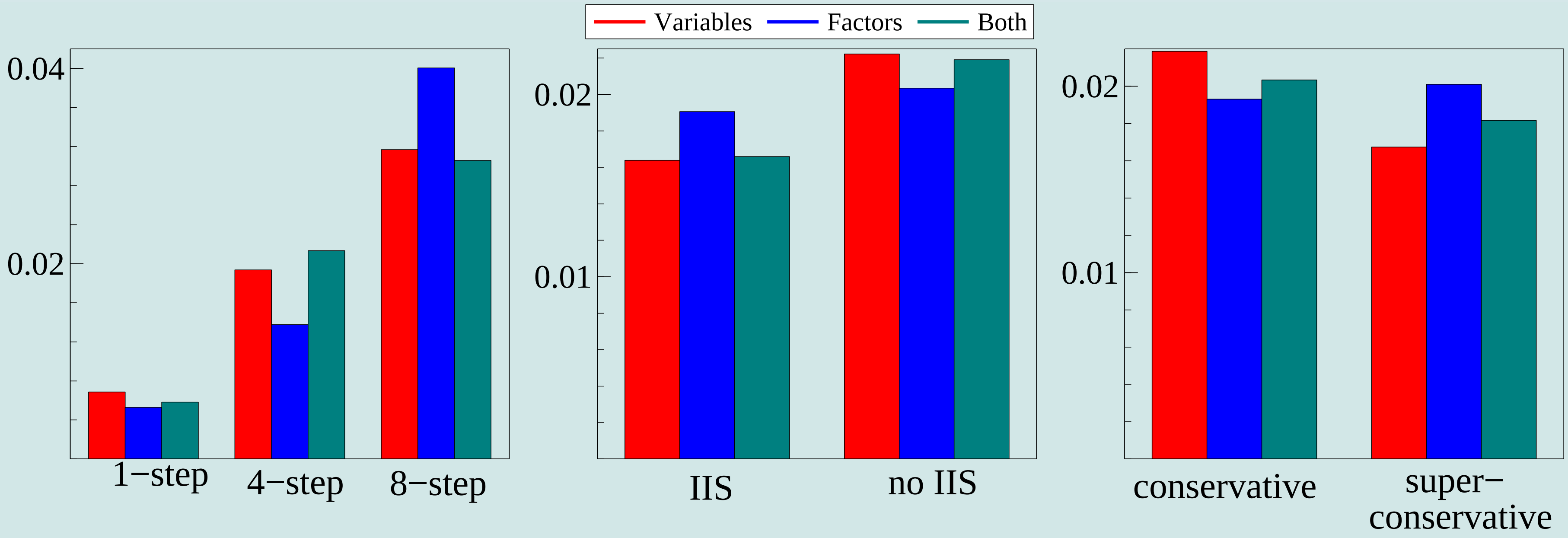
$\widehat{\sigma} =$ equation standard error, No. regressors and No. dummies are number of regressors and, as a subset, number of dummies retained: Cons and Super are conservative and super-conservative strategies.

RMSFE averaged across horizon, IIS or not, and α for GDP and GDP growth.

	Variables	Factors	Both	RW	AR(D)	AR(I)
$\Delta\widehat{y}_{T+k}$	0.666	0.600	0.650	0.666	0.485	0.491
$\widehat{y}_{T+k}$	1.931	1.971	1.925	3.509	1.324	1.336

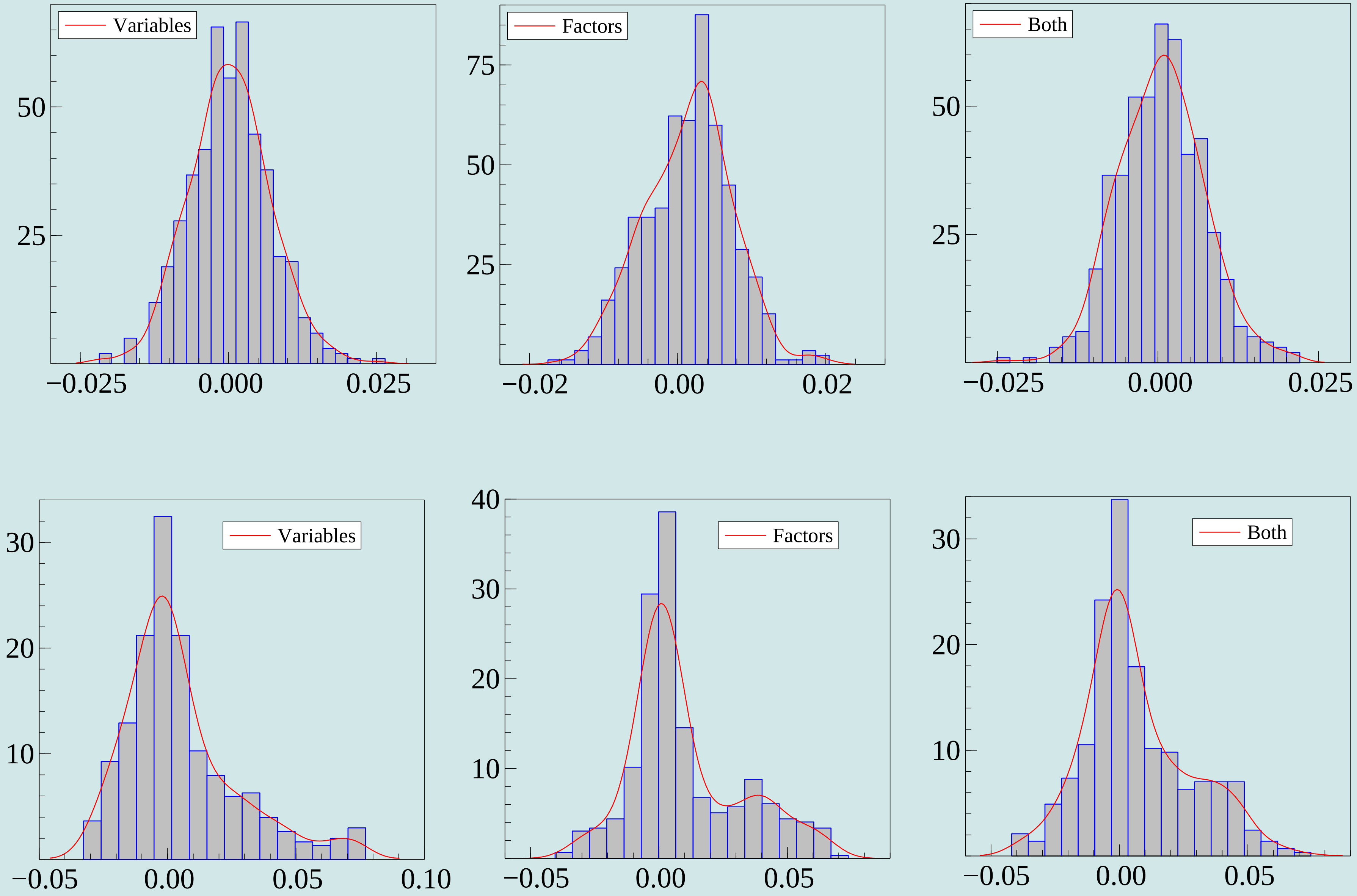


Average RMSFE for log GDP $\left(\widehat{y}_{T+h}\right)$



Factor model best at short horizons but worst at long.
Variables model improves in forecast accuracy at 8-step.
Some benefit to IIS but few indicators retained at tight α (< 6), roughly the same for factors and variables models.
In levels, factor forecasts worsen as horizon increases.
IIS yields greater improvements as retained dummies translate to level shifts in 1984.

Distributions of forecast errors



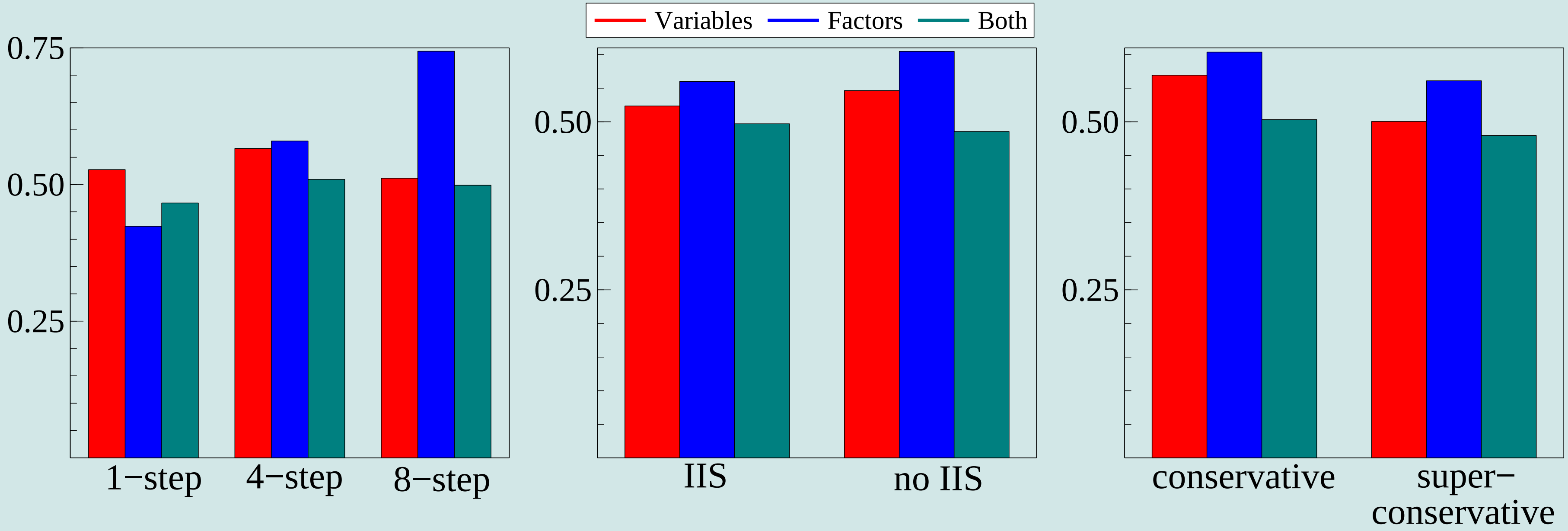
In-sample results for inflation models with IIS

	1-step		4-step		8-step	
	Cons	Super	Cons	Super	Cons	Super
Variables						
$\widehat{\sigma}$	0.240	0.283	0.293	0.290	0.227	0.296
No. regressors	20	12	17	13	21	12
No. dummies	2	1	4	8	8	3
Factors						
$\widehat{\sigma}$	0.263	0.336	0.299	0.318	0.311	0.327
No. regressors	15	7	20	14	21	20
No. dummies	5	1	4	13	7	6
Both						
$\widehat{\sigma}$	0.266	0.284	0.302	0.356	0.284	0.334
No. regressors	14	11	12	10	13	8
No. dummies	4	3	4	3	1	3

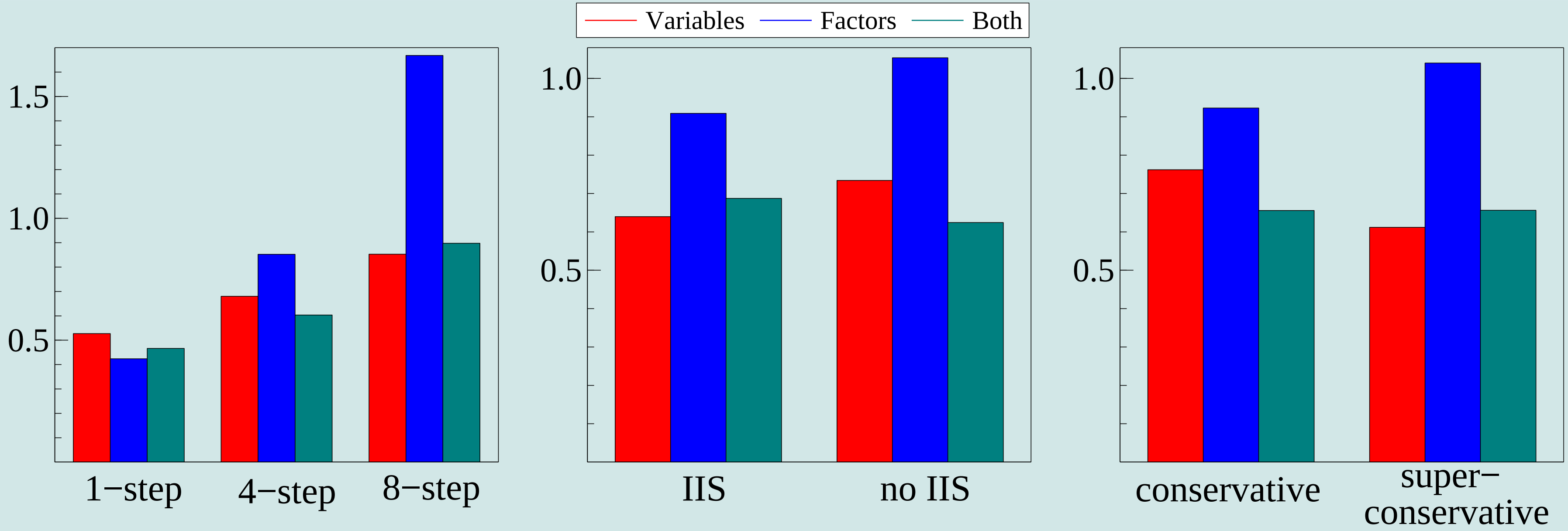
$\widehat{\sigma} =$ equation standard error, No. regressors and No. dummies are number of regressors and, as a subset, number of dummies retained: Cons and Super are conservative and super-conservative strategies.

RMSFE averaged across horizon, IIS or not, and α for inflation and the change in inflation.

	Variables	Factors	Both	RW	AR(D)	AR(I)
$\Delta\widehat{y}_{T+k}$	0.535	0.582	0.491	0.577	0.416	0.415
$\widehat{y}_{T+k}$	0.687	0.981	0.656	0.469	0.444	0.464

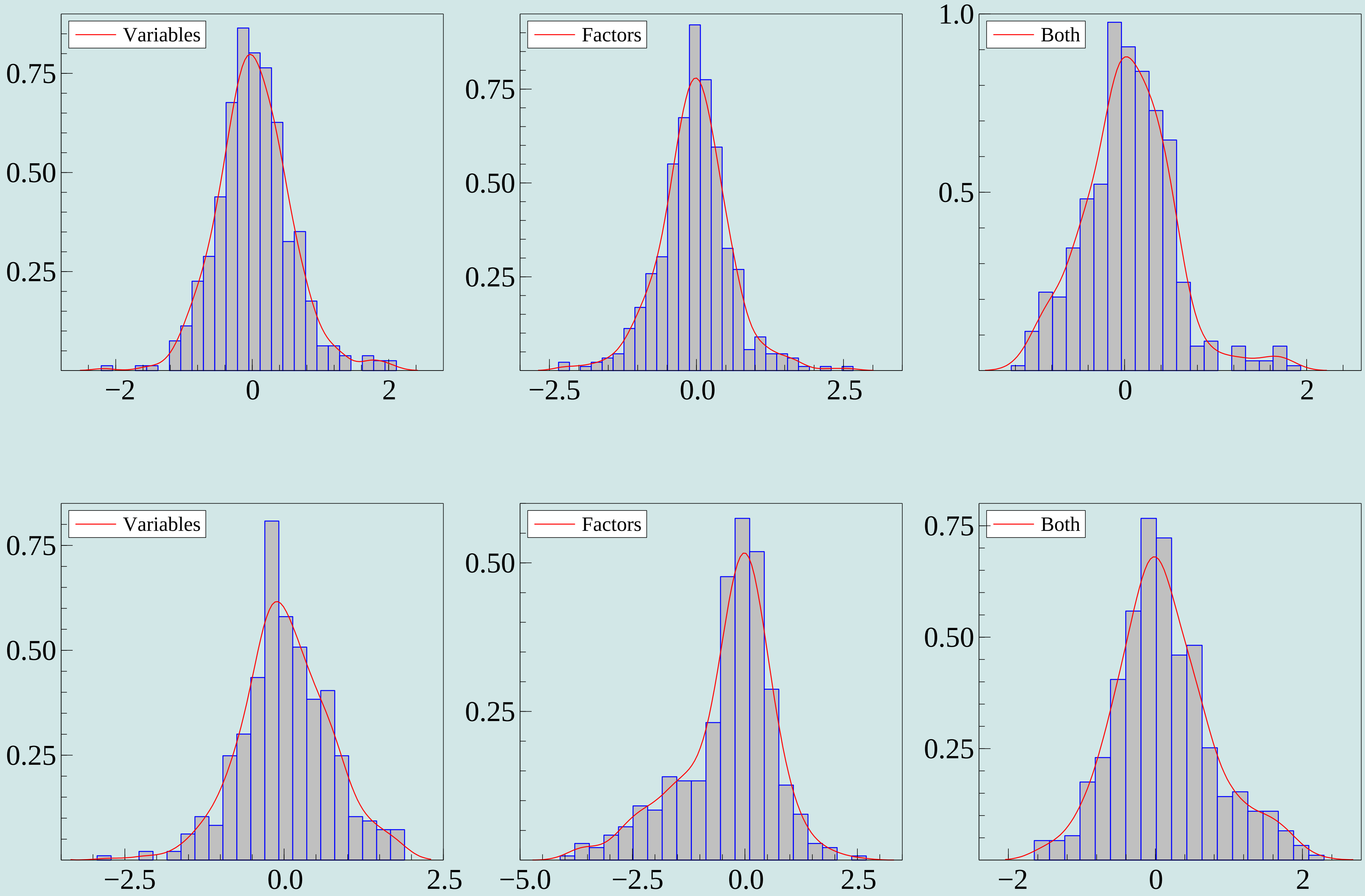


Average RMSFE for inflation $(\widehat{y}_{T+h})$



All models similar over forecast horizon
but factor models worst on average as horizon increases.
Small gains to IIS for variables or factor models.
Tighter α improves forecasts,
parsimonious models yield smaller RMSFE.
Factor models most useful for shorter horizons.
Levels highlight worse forecast accuracy of factors.

Distributions of forecast errors

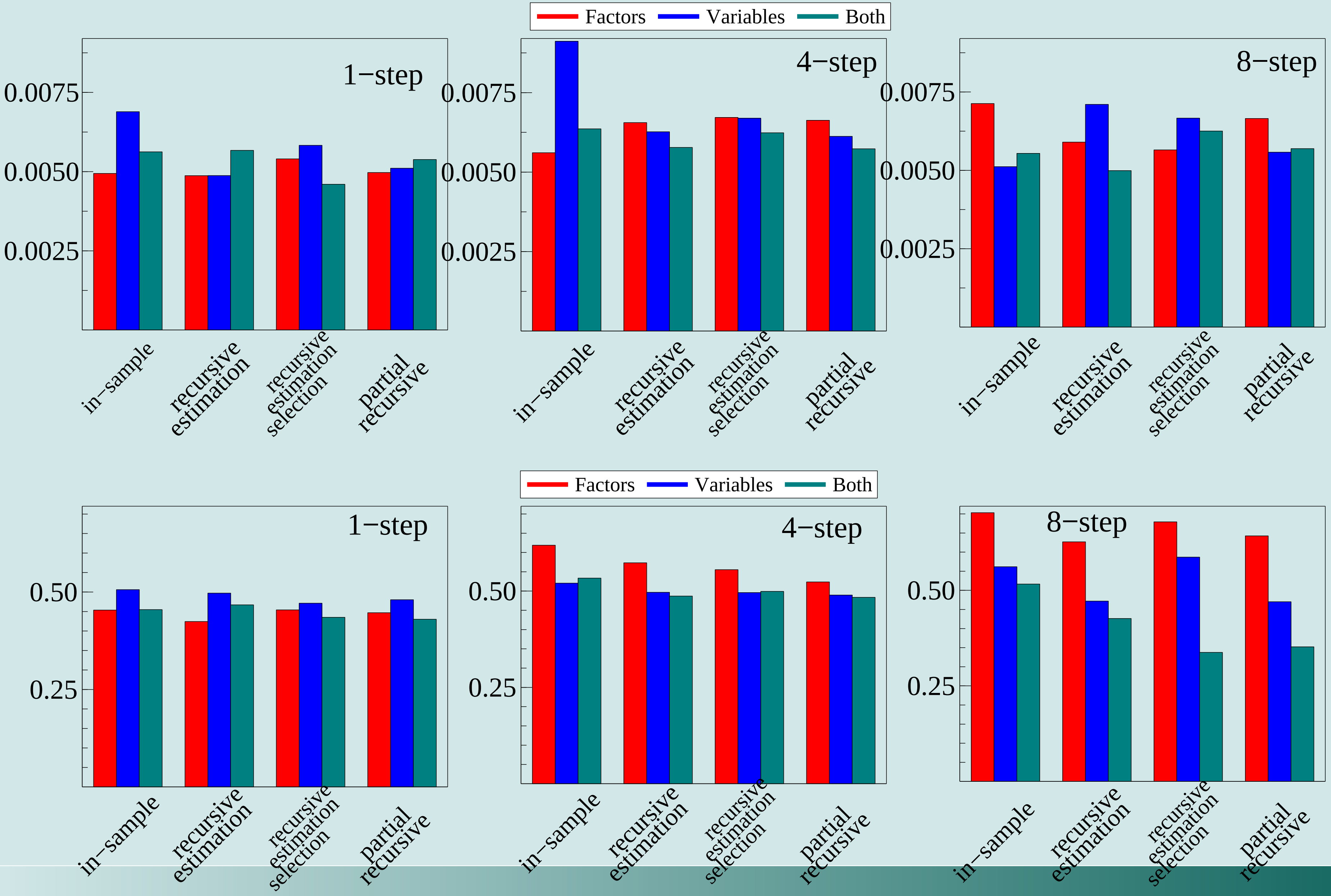


Selection and estimation method

Does selection and estimation affect forecast accuracy? Compare four methods:

1. model is fixed after selection in-sample over $t = 1, \ldots, T$;
2. model is selected over $t = 1, \ldots, T$, parameters and eigenvalues recursively re-estimated over each forecast horizon, $T = 1, \ldots, T + h$;
3. model is both selected and re-estimated recursively over each forecast horizon, $t = 1, \ldots, T + h$, with eigenvalues re-calculated, so specification can change with each new observation; and
4. partial recursive selection: previously retained variables are not dropped even if insignificant and selection is only over additional variables or factors.

RMSFES for selection and estimation method



Route map

(1) **Setting the scene**

(2) **Relating variables $\{\mathbf{z}_t\}$ to factors $\{\mathbf{f}_t\}$**

(3) **Comparing variable-based and factor-based models**

(4) **Taxonomy of forecast errors in factor models**

(5) **Forecast failure in equilibrium-correction models**

(6) **Model selection with both factors and variables**

(7) **Forecasting US GDP and inflation**

Conclusions

Conclusions

Can jointly include principal components and variables.
Automatic selection allows $N > T$, perfect collinearities and multiple breaks.
Autometrics can select models from both PCs and variables.
Variables models and factor models are EqCMs, so non-robust to location shifts.
Impulse-indicator saturation can account for shifts in-sample, and could be used for intercept corrections at forecast origin.
Find some advantage to using IIS even for forecasting in differences: levels forecasts highlight the level shifts.
Empirical results for GDP and inflation are mixed: factor models seem more useful for 1-step forecasts.

Conclusions II

For multi-step forecasting, *Autometrics* selection over variables and PCs better than factor forecasts:
benefits to selecting weights from correlations with y_{t+h}.
Little evidence that recursive estimation or recursive estimation and selection is better.
Gains to using IIS at tight α controlling null retention rate.

Selection jointly on factors and variables avoids imposing specific form, circumvents arbitrary choice of factors.

Whether data better described by factors or variables depends on phenomena,
but can be determined by data using model selection techniques with low cost from retaining irrelevant effects.

References

Box, G. E. P., and Jenkins, G. M. (1970). *Time Series Analysis*. San Francisco: Holden-Day.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2011a). Model selection for multiple breaks. *JoE*.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2011b). Forecasting Breaks and During Breaks. In Clements, M. P., and Hendry, D. F. (eds.), *Oxford Handbook*, pp. 315–353. OUP.

Castle, J. L., and Shephard, N. (eds.)(2009). *Methodology & Practice of Econometrics*. OUP.

Clements, M. P., and Hendry, D. F. (2005). Information in forecasting. *OBull*, **67**, 713–753.

Doornik, J. A. (2009). Autometrics. In Castle, and Shephard (2009), pp. 88–121.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized factor model: Identification and estimation. *Review of Economics and Statistics*, **82**, 540–554.

Hendry, D. F. (2006). Robustifying forecasts from EqCMs. *JoE*, **135**, 399–426.

Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. ONR Research Memorandum 52, Carnegie Institute, Pittsburgh, Pennsylvania.

Johansen, S., and Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In Castle, and Shephard (2009), pp. 1–36.

Klein, L. R. (1950). *Economic Fluctuations in the United States, 1921–41*. No. 11 in Cowles Commission Monograph.

Makridakis, S., and Hibon, M. (2000). The M3-competition. *IJF*, **16**, 451–476.

Persons, W. M. (1924). *The Problem of Business Forecasting*. No. 6 in Pollak Foundation.

Smets, F., and Wouters, R. (2003). An estimated stochastic dynamic general equilibrium model of the Euro Area. *JEEA*, **1**, 1123–1175.

Stock, J. H., and Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macro-Economic Annual*, 351–409.

—— (1999). Linear and nonlinear models for forecasting macroeconomic time series. In Engle, R. F., and White, H. (eds.), *Cointegration, Causality and Forecasting*, pp. 1–44. OUP.

—— (2009). Forecasting in dynamic factor models. In Castle, and Shephard (2009), Ch. 7.

Wallis, K. F. (1989). Macroeconomic forecasting: A survey. *Economic Journal*, **99**, 28–61.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, **6**, 324–342.

Zarnowitz, V., and Boschan, C. (1977). Cyclical indicators. In *Handbook of Cyclical Indicators*, pp. 170–183. Washington, USA: U.S. Department of Commerce.

Retracing route

(1) Setting the scene

(2) Relating variables, $\{\mathbf{z}_t\}$, to factors $\{\mathbf{f}_t\}$

(3) Comparing variable-based and factor-based models

(4) Taxonomy of forecast errors in factor model

(5) Forecast failure in equilibrium-correction models

(6) Model selection with both factors and variables

(7) Forecasting US GDP and inflation

Conclusion

Impulse-indicator saturation

To tackle multiple breaks & data contamination (outliers), add T impulse indicators to candidates for T observations.

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

μ is parameter of interest

Uncertain of outliers, so add T indicators $\mathbf{1}_{\{t=t_i\}}$

to set of candidate regressors.

First, include half of indicators, record significant:

just 'dummying out' $T/2$ observations for estimating μ

Then omit, include other half, record again.

Combine sub-sample indicators, & select significant.

αT indicators selected on average at significance level α

Feasible 'split-sample' impulse-indicator saturation (IIS) algorithm

Dynamic generalizations

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 $\boldsymbol{\beta}$ (not selected) and second moment $\boldsymbol{\Sigma}$:

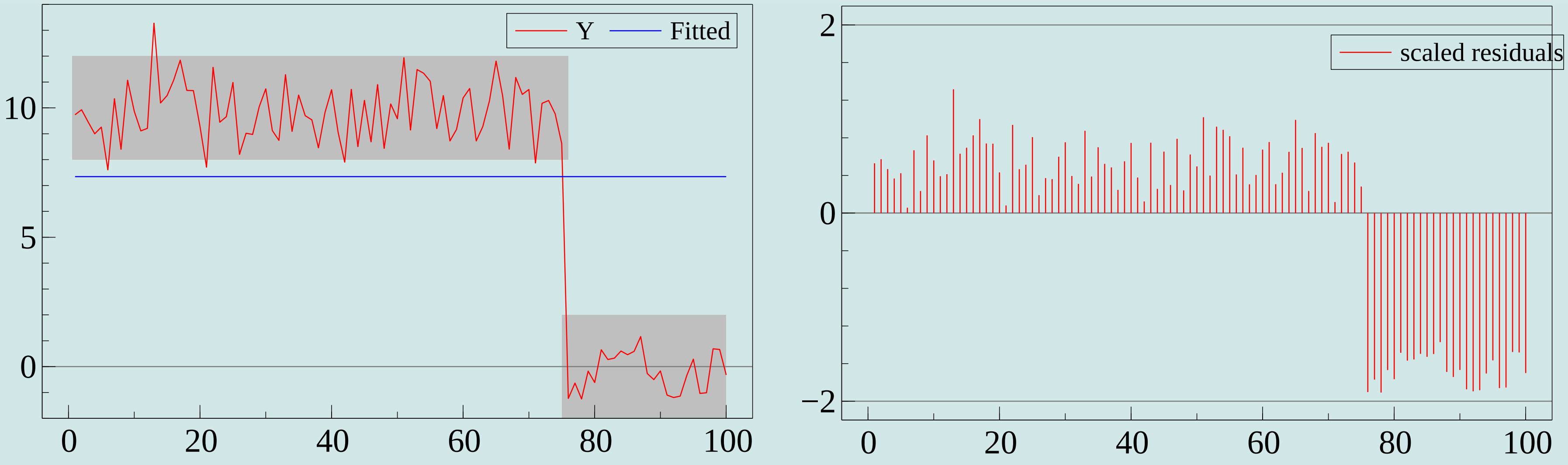
$$T^{1/2}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{\mathsf{D}}{\rightarrow} \mathsf{N}_n \left[\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{\Sigma}^{-1} \boldsymbol{\Omega}_\beta\right]$$

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

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

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

Structural break example

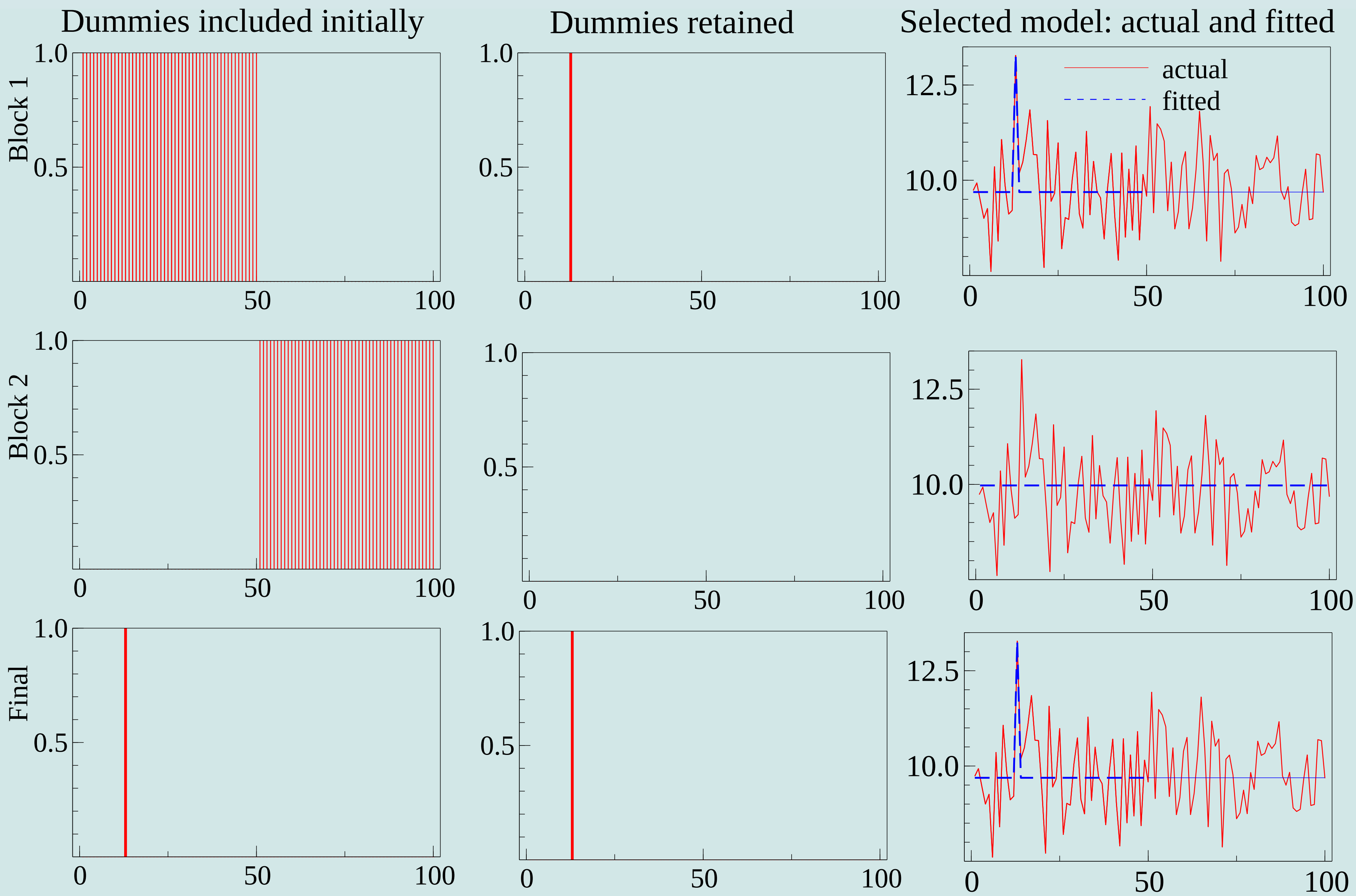


- Size of the break is **10 standard errors** at $0.75T$
- There are **no outliers** in this mis-specified model as all residuals $\in [-2, 2]$ SDs:

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

Let's see what Autometrics reports

Null 'split-sample' search in IIS



‘Split-sample’ search in IIS